2008

CONDUCTING SYSTEMATIC REVIEWS FOR SOCIAL POLICY: THE ROLE OF UNDERSTANDING DISCOURSES IN METHODOLOGICAL DEVELOPMENT

PEARSON, MARK

http://hdl.handle.net/10026.1/1818

http://dx.doi.org/10.24382/4936
University of Plymouth

CONDUCTING SYSTEMATIC REVIEWS FOR SOCIAL POLICY: THE ROLE OF UNDERSTANDING DISCOURSES IN METHODOLOGICAL DEVELOPMENT

by

MARK PEARSON

A thesis submitted to the University of Plymouth
in partial fulfilment for the degree of

DOCTOR OF PHILOSOPHY

School of Law & Social Science
Faculty of Social Science & Business

In collaboration with the
National Institute for Health and Clinical Excellence (NICE)

February 2008

Conducting Systematic Reviews for Social Policy: The Role of Understanding Discourses in Methodological Development

Mark Pearson

For Nicole.

Thankyou.

Conducting Systematic Reviews for Social Policy: The Role of Understanding Discourses in Methodological Development

Mark Pearson

Abstract

This thesis uses a case study of the process of conducting a systematic review in the field of substance misuse in order to analyse critically how knowledge is cumulated for the purposes of informing social policy. The analysis is grounded in two areas of social research that are seldom drawn together; the methodological (in this instance, the work of the social research methodologist Donald Campbell) and the linguistic (in the form of a measured application of discourse analysis). By means of this dual approach it is proposed that a better understanding can be obtained not only of how systematic review methods may be usefully developed, but also of the substantive impact that the way in which those methods are discussed and debated (through discourse) can have upon the development. In this way, the process of conducting a systematic review for social policy is critically located within an understanding of both policy making and methodological development as discursive processes. This is important for the way that it allows evidence for policy and practice (both in the sense of the framing of the evidence and the methods used to synthesize it) to be discussed in terms that prioritize respectful debate rather than the promotion of particular methods as superior for the production and synthesis of knowledge. Furthermore, it enables a critical understanding of how dominant discourses can not only frame policy issues, but also the production of evidence-bases that are subsequently used in the policy making process.

Contents

Figures and Tables

Abbreviations

CASE	- Collaborative Awards in Science and Engineering (studentship)
EBM	- Evidence-Based Medicine
EBP	- Evidence-Based Policy and Practice
EM	- E-Mail – e.g. (EM 04/08/06 3-5) refers to lines 3-5 of an e-mail sent on August 4th 2006
ESRC	- Economic and Social Research Council
FBN	- Federal Bureau of Narcotics
HDA	- Health Development Agency
ID	- Interview number – e.g. (ID1 4-8) refers to interview 1, lines 4-8
LREC	- Local Research Ethics Committee
LST	- Life Skills Training programme
MT	- Managing Team (of the case study systematic review)
NHS	- National Health Service
NICE	- National Institute for Health and Clinical Excellence
RCT	- Randomized-Controlled Trial
RD	- Research Diary – e.g. (RD 10-15 06/07/06) refers to lines 10-15 of the research diary entry of July 6th 2006
RT	- Review Team (for the case study systematic review)
SR	- Systematic Review
SSOT	- Society for the Suppression of the Opium Trade
UK	- United Kingdom
US	- United States (of America)

Acknowledgements

Thanks are due to Dr Ross Coomber (University of Plymouth), Professor Mike Kelly (National Institute for Health and Clinical Excellence), and Professor Malcolm Williams (University of Plymouth) for supervising the conduct of this research. Staff at the National Institute for Health and Clinical Excellence (and at the National Collaborating Centre for Drug Prevention, Liverpool John Moores University) are also due my thanks for their assistance in facilitating the conduct of this research, as well as becoming willing participants in it. In particular, Simon Ellis and Nichole Taske tolerated my intrusions with good humour and went the extra mile to provide the privileged access that enabled the research to be completed successfully. In this respect, Mike Kelly should be acknowledged a second time for his guiding hand in managing the conduct of the research at the coalface of knowledge synthesis.

The support of other postgraduates is vital for climbing the 'PhD mountain', and I'd like to single out two who have now successfully gained their doctorates; Sue Richardson (University of Plymouth) and Sara Shaw (University College London). Both have acted as non-judgmental mentors to me and have provided valuable feedback that has been incredibly helpful for finding my way through the research process. An honourable mention here too for Steve Parkin (University of Plymouth) for his insights into substance misuse research and Newcastle United FC.

Finally, thanks are due to the ESRC for funding the CASE studentship (in partnership with NICE) under which this research was conducted.

Author's Declaration

At no time during the registration for the degree of Doctor of Philosophy has the author been registered for any other University award without prior agreement of the Graduate Committee.

This study was financed with the aid of an Economic & Social Research Council Collaborative Awards for Science and Engineering (CASE) postgraduate studentship (award number PTA 033 2004 00074) and carried out in collaboration with the National Institute for Health and Clinical Excellence (NICE).

A programme of advanced study was undertaken and aspects of the research presented at conferences and through peer-reviewed publications, as detailed below.

Publications arising from this work:

Pearson, M. (2008, forthcoming) Evidence-Based Policy & Practice: How can research be made more relevant to practitioners and policy-makers? In: M.Wright (ed.), *Proceedings of the Faculty of Social Science & Business (University of Plymouth) Postgraduate Symposium 2007*. Plymouth: University of Plymouth

Pearson, M. (2007) Systematic Reviews in Social Policy: To go forward, do we first need to look back? *Evidence & Policy* 3 (4) 505-526

Pearson, M. (2006) Evidence in social policy – can we see past the hierarchy of evidence? In: A.Barton and J.Lean, (eds.), *Proceedings of the School of Sociology, Politics & Law and Plymouth Business School (University of Plymouth) Postgraduate Symposium, January 2006*. Plymouth: University of Plymouth

Presentations given:

'Systematic Reviews in Social Policy: To go forward, do we first need to look back?' – School of Law & Social Science (University of Plymouth) seminar series (October 11th 2007)

'Evidence-Based Policy & Practice: How can research be made more relevant to health policy makers?' – British Sociological Association Medical Sociology conference (September 7th 2007)

'Evidence-Based Policy & Practice: Do Systematic Reviews permit valid inferences to be made?' - London School of Economics Eighth Inter-University Graduate conference (May 19th 2007)

'Evidence-Based Policy & Practice: How can research be made more relevant to practitioners and policy-makers?' - Faculty of Social Science & Business (University of Plymouth) Postgraduate Symposium (March 3rd 2007)

'Evidence in Social Policy: Can We See Past the Hierarchy of Evidence?' - Plymouth Business School and the School of Sociology, Politics, and Law (University of Plymouth) Postgraduate Symposium (January 14th 2006)

Conferences attended:

Enquiry, Evidence and Facts: An Interdisciplinary Conference, The British Academy (London) (December 13th-14th 2007)

Randomized Controlled Trials in the Social Sciences: The Way Forward, Second Annual Conference, York Trials Unit (University of York) (September 13th-14th 2007)

United Kingdom Public Health Association 15th Annual Conference (Edinburgh) (March 28th-29th 2007)

Methodological Innovations Methods Event, University of Plymouth (April 19th-20th 2007)

The Seventh Annual International Campbell Collaboration Colloquium (London) (May 15th-16th 2007)

British Sociological Association Annual Conference (Harrogate) (April 14th-16th 2006)

United Kingdom Public Health Association 14th Annual Conference (Telford) (March 14th-15th 2006)

British Sociological Association Medical Sociology conference (York) (September 8th-10th 2005)

Word count of main body of thesis: **76,867**

Signed.....

Date....20/05/2008

Chapter 1

Knowledge Construction and Evidence-Based Policy and Practice

The end of the twentieth century was notable for a move to greater openness in the United Kingdom's (UK)[1] government towards 'evidence' that could inform policy and practice. Whilst it was indeed the case that a key New Labour election pledge was 'what matters is what works', and that formal commitment to this 'modernization' agenda was made across policy sectors (Cabinet Office, 1999), the move towards the greater utilization of evidence in the policy making process had arguably been growing before New Labour's crystallization of the approach. An increasingly well-informed public were demanding that publicly-funded services were effective; in addition there was growing public distrust in professional expertise (linked with the rise in consumer power) and the proliferation of non- and quasi-governmental bodies (pressure groups, 'think-tanks', professional bodies, watchdogs, and regulators) who were all seeking to marshal evidence in order to influence policy making (Davies et al., 2000; Solesbury, 2001). However, this substantial growth in desire to utilize evidence for the public good was not matched by the development of methods for bringing together disparate forms of evidence in a rigorous fashion:

> "...on virtually any [social science] topic you can name, there is a vast body of past research that may have some continuing value but mostly remains ignored. Social science is very bad at the cumulation and re-use of past research results." (Solesbury, 2001, p.5)

This thesis investigates an example of knowledge cumulation in the social sciences through the analysis of the process of constructing an evidence-base in the field of substance misuse. The thesis starts out with a wide perspective on the production of knowledge, tightens this perspective in order to relate specifically to systematic reviews and the historical construction of drug policy in the UK, and then

[1] In order to allow the chapters in this thesis to be read individually (without constantly referring back to the list of abbreviations) the first occurrence in each chapter of each abbreviation is also provided in full.

focuses tightly upon the analysis of the case study. The perspective is then widened again so as to integrate the analysis with the issues surrounding epistemology, ontology, methods, and the framing of drug policy (contained in the first three chapters) in order to consider how the analysis can contribute to the development of SR methods as a whole.

Chapter 1 provides the context for the process of knowledge production that is analysed in the case study. The philosophical basis for the vital roles played by theory and values in the conduct of scientific research is set out, and the case for considering research to be an inherently social process is made. The importance of understanding ontology with regard to the production of valid knowledge is then illustrated by a consideration of idealism and realism, and the implications of a realist ontology for evaluating social policy described. With this philosophical background in place, the socio-political context in which Evidence-Based Medicine (EBM) and Evidence-Based Policy and Practice (EBP) developed is examined with regard to the development of the field of public health and the rise of New Public Management (NPM). This provides the context for an analysis of the development of EBM in the United Kingdom (UK) and the lessons that may be drawn from it regarding the development of systematic review (SR) methods for social policy. Whilst the policy making process is acknowledged to be far more complex than the straightforward implementation of research evidence, the case for there being an important, indeed vital, role for this evidence is made.

Chapter 2 focuses on SR methods in order to better understand the bases upon which a wide spectrum of methods are advocated and the reasons why advocates of different methods do not always engage constructively. Donald Campbell's work on social research methods is introduced as a backdrop to the contemporary debate over SR methods. The case is made for the potential this work has for contributing to the

development of SR methods, most importantly with regard to the concepts of internal, external, and construct validity.

Chapter 3 details the history of drug policy making in the UK in order to place the topic of the case study SR (preventing and reducing substance misuse in vulnerable and disadvantaged youth) within its historical and social context. Whilst the focus is on the development of UK drug policy, this is set within the wider international context so that the inter-relationship between UK and international policy may be better understood. An analysis of discourses that have informed drug policy making from the mid-nineteenth to early twenty-first century is presented in order to inform the analysis of the case study review's recommendations.

Chapter 4 details the rationale for utilizing a case study in the research conducted for this thesis. The epistemological and ontological positions taken in the conduct of the research are set out, and the rationale upon which a balance between internal, external, and construct validity was struck is provided. The reasons for utilizing participant-observation, interviews, and documentary analysis as research methods are set out, as is the manner in which ethical research practice was attained.

Chapters 5 to 7 contain the analysis of the case study. Chapter 5 focuses upon the task of producing the SR itself and the issues.that arose in doing so, such as the relative allocation of time to different tasks in the review and ambivalence about the role of judgement and expertise. The SR database produced by the review team is also utilized in order to analyse a purposive sample of research papers that were excluded from the SR in order to explore what these papers might have been able to contribute to the review.

Chapter 6 focuses upon the utilization of the SR by a committee in order to make recommendations for policy. The issues that arose from this process are analysed with regard to the way in which evidence was selected, weighted, and synthesized in the form of evidence statements made in the SR. The benefit of hindsight is used to investigate how evidence from a purposive sample of included studies might have been utilized in different ways that could have better informed the committee. The manner in which discourses about SRs and substance misuse governed the deliberations of the committee is also analysed, and the implications for the focus of the final recommendations are discussed.

The analysis of case study material presented in Chapter 7 is concerned with developing the grounds for the development of SR methods. The reflections of senior members of the review's managing team are analysed within a framework that melds the work of Donald Campbell with the insights into both topic and methods attained through discourse analysis. The concluding chapter (8) critically reflects upon the research methods used in this thesis, and draws together the strands of epistemology, ontology, methods, and the framing of drug policy by considering how Campbell's concepts of research validity could usefully inform the development of SR methods for social policy.

1.0 The construction of knowledge

Knowledge is not simply collected, but instead has to be actively constructed. This process of construction can be rigorous to a greater or lesser degree. To better understand how knowledge is constructed, this chapter begins by considering how philosophers of science have approached the issue. The roles played by theory, the manner in which argumentation is utilized, and the possibility or otherwise of value freedom are considered regarding their impact upon knowledge construction. The

understanding of different ontologies is argued to be a key way in which different approaches to knowledge construction can be better understood.

1.1 The framing of observations by theory

The empiricist belief that objective scientific endeavour was best served by observing, and in essence simply allowing the real properties of objects to be perceived by an open mind, was comprehensively critiqued by Karl Popper. For Popper (1972, p.46-48), the description of complex phenomena required observation that was selective, for a descriptive language was necessarily grounded in a particular classificatory system that is in itself grounded in a particular framing of the world. Popper was nevertheless clear that science remained a rational affair; indeed, the theory-laden nature of observation provided better grounds for science to proceed (by mutual debate) than the authoritarian approach that was engendered by empiricism. Popper argued that the empiricist view, "... that truth is manifest... [and] is there for everyone to see, if only he wants to see it" (Popper, 1972, p.8), led to a dogmatic approach to the construction of knowledge that risked becoming highly arbitrary[2].

As noted by Popper, there is a cognitive dimension to observation, whereby the observer utilizes a classificatory system as a means of attaining a grasp of the phenomena observed. It is thus possible to see that two observers, using different classificatory systems, may see different things in the same phenomenon (Williams and May, 1996). Furthermore, there is a social dimension to observation; if it is agreed that theory guides not only what objects of knowledge are observed in the first place, but also the aspects of those objects that are perceived and recorded, then it is impossible for theories to be constructed outside of a social system. In short, there is no 'neutral' starting point, although we may indeed argue that there are stronger

[2] However, it should be noted that Popper (1972, p.5-9) acknowledge that this 'epistemological optimism' (that truth could be clearly distinguished from falsehood if the appropriate methods were used) underlay the Enlightenment and thereby lay the foundations for societies based upon reason rather than superstition. For Popper, it was a case of "... a bad idea inspiring many good ones." (Popper, 1972, p.8).

grounds for utilizing certain theories over others. To observe, as opposed to just passively 'experience', requires a theory to make those observations intelligible (Williams and May, 1996).

In the philosophy of science, different perspectives exist on the manner in which theory is utilized in the course of scientific research. For Goodman (1978), the nature of research is such that the observations made can only, even at best, provide a modest basis for the construction of knowledge. It is left to the researcher to 'fill in' the substantive gaps, which necessarily reflect a theoretical position. A more positive argument is expressed by Lipton (2004, p.55-61), where the researcher is not so much 'filling in' the gaps as utilizing theory in an iterative process that works towards 'inferences to the best explanation'. In this view, the conduct of research cannot be neatly delineated into stages of inference and explanation. Rather, theory provides the framework within which explanation is worked towards, and the inferences that are made in the course of conducting research are made with these explanatory considerations being borne strongly in mind (Lipton, 2004, p.89-90).

The effect of the framing of observations by theory has been most cogently critiqued by the feminist philosopher of science, Helen Longino. Longino is mindful of the manner in which the social and cultural values of a patriarchal society inform the theories which scientists utilize to observe the world, but nevertheless contends that a 'neutral' theory would result in descriptions of the world that are 'impoverished' (Longino, 1990, p. 40-48 and 219-225). What concerns Longino is that the values which underlie theory (and hence frame observations) remain unacknowledged and hence unexamined. In this way, theory-laden observations that reflect a particular world view (for Longino, a patriarchal world view) come to be accepted as neutral and objective simply because their partiality remains invisible (Longino, 1996). This 'invisibility' is compounded by the nature of being a member of a scientific community

where membership is in many ways conditional upon *not* reflecting upon this issue (Longino, 1996).

1.2 Limitations of using deductive and inductive arguments

Using arguments is an inherent part of scientific endeavour; at one end of the process, the rationale for initiating a research project will need to be convincingly stated in order to obtain funding, whilst at the other, arguments will need to be used in order to defend the research against the critique of the scientific community. This section focuses upon just one key aspect of the use of argument within the research process; the utility of deductive and inductive arguments as a means of progressing from the particularities of research findings to generalizations (in certain instances, laws) that are more widely applicable. In doing so, it is argued that although there are sound reasons for preferring inductive over deductive arguments, it is still not the case that an argument can ever be 'proven' to be correct.

The conclusion of a deductive argument follows logically from its premises. This would be a tenable position to support in the scientific process if it was the case that our linguistic expressions corresponded directly with the 'reality' of the phenomena they were intended to represent. However, deductive arguments are themselves founded on concepts that are arrived at by inductive arguments, meaning that only relationships between concepts can be specified as opposed to relationships that 'actually' exist in the real world (Williams and May, 1996, p.25-32). The Vienna Circle of logical positivists held this problem to be insoluble, and driven by a desire to provide the scientific process with a tangible foundation, elected to pursue the phenomenalist approach whereby the observation of a phenomenon alone would verify its existence. Popper's falsificationist thesis strongly critiqued the Vienna Circle on the grounds that phenomenalism led to a verificationist approach that had a tendency to produce confirming instances alone, rather than subjecting theories to the strongest critique (Popper, 1972).

Inductive arguments work from specific instances to a general conclusion, and as such go beyond the premises contained in the statement concerned (Chalmers, 1999). Arguably, making these inferences is an inherent part of any scientific statement that is proposed to have relevance to more than the specific phenomena observed at a particular time and space. The manner in which inductive arguments make statements that go beyond their premises means that that a scientific theory can never be 'proven' in the sense of being logically deduced from the facts. Despite the strengths of inductive arguments in making rigorous inferences that have utility for decision-making in the real world, this does not mean that inductive arguments are entirely adequate for the task at hand (deriving scientific knowledge from the facts by induction) as all such arguments are themselves grounded in inductive knowledge. Chalmers (1999, p.48) refers to this as a "never-ending chain", where inductive arguments appeal to prior knowledge, which itself is justified by inductive arguments that are grounded in prior knowledge, and so on, *ad infinitum*.

1.3 Value freedom and the quest for 'objectivity'

'Objectivity' is a prized goal in the conduct of science and may even be argued to be the foremost criterion that differentiates scientific knowledge from other particularistic and 'subjective' sources of knowledge. Objectivity is premised upon the subject-object distinction (value freedom), the notion that the subject (researcher) can only attain objectivity by being separate from the object (the phenomenon observed). Moreover, in this view, the subject's values cannot impinge upon the observation of the object, for to do so would produce knowledge that could not be differentiated from non-scientific sources (Bunge, 1998). Whilst not dismissing the pursuit of reliable knowledge, Nagel (1979) has contended that such an approach is unattainable in view of the multitude of assumptions contained within a value system and which underlay the conduct of scientific research. Is it possible for science to proceed rationally whilst these values remain implicit and unexamined?

A strong case may be made that scientific research of both the natural and social worlds are thoroughly, and by no means disastrously, imbued with values. These values exist along a continuum, from 'values' in the numeric sense of the word to 'values' in the social sense (Williams, 2005, p.105-106). What links these apparently disparate definitions of values is the interdependency that exists between them; a measurement scale may be objective and reliable, but social values are involved in making the decision to utilize a particular scale; similarly, social values can play a key role in shaping the topics deemed worthy of scientific investigation and the manner in which those topics are approached (Williams, 2005, p.112-114). Values, including the value of objectivity itself, cannot and should not be partitioned off from the conduct of science. To do so removes an important and inherent part of the scientific process that, far from increasing the rigour and reliability of research, actually impoverishes the pursuit of knowledge.

1.4 Can socially-constructed knowledge be rigorous?

Scientific knowledge has been argued to be distinctive from other forms of enquiry by its explanatory power; in short, to rigorously construct knowledge about the nature of relationships between phenomena that may, superficially, appear unrelated (Nagel, 1979). Moreover, the co-ordinated nature of scientific inquiry is proposed to ensure that such knowledge is not only open to informed criticism by the scientific community, but also integrated into a cumulative body of knowledge that provides the basis for future scientific investigation (Popper, 1972). Clearly, there may not be agreement within a scientific community upon whether a 'piece' of knowledge is sufficiently rigorous for it to be incorporated, for different views are held regarding what constitutes valid knowledge. Disagreements such as this highlight the social nature of science, for claims and counter-claims are made within a socio-cultural context. This is vital to acknowledge if scientific research is not to be elevated to a position beyond critique; science cannot claim privileged access to the truth, but it can justifiably report

conclusions as a 'critically achieved consensus' of a scientific community (Longino, 1990, p.78-81). For example, Popper's falisificationist thesis rested upon the nature of debate within scientific communities; if there was no forum in which knowledge could be discussed and critiqued, the possibility of advancing knowledge through falsification was slim (Popper, 1972, p.33-37 and 50-52). Nagel (1979, p.489) similarly counselled that "... mutual exchange... [and] the free but responsible criticism of ideas" were essential to the advancement of scientific knowledge.

The social nature of knowledge does not mean that it is constructed in the relativistic sense of existing only because members of a community agree that it is so; rather, it is constructed in the sense of a designer making the best use of the materials available to produce a meaningful whole. This construction of knowledge is grounded in the social world and the values that are dominant. It is for this reason that Longino (1990; Longino, 1996) argues for scientific knowledge to be open to critique and debate outside of the scientific community from which it originated; to not do so is to allow the values that underlay the construction of knowledge to be ossified in the form of 'objectivity' and hence placed beyond the reach of criticism outside of the scientific community.

The scientific process, properly applied, has been argued by many to be one that is characterized by humility; researchers should be willing to revise theories in the light of knowledge and to acknowledge and address errors in research where they occur (Popper, 1972; Oakley, 2000). Few would argue against such a position, but it does tend towards a conceptualization of knowledge that is produced outside of a social system. If Longino's (1990; 1996) thesis is correct (regarding the way in which values substantively permeate research programmes that are ostensibly wholly objective), how might rigorously constructed knowledge be attained in a way that does not simply factor out the role played by values? Williams (2005, p.108-110) suggests

that a 'situated objectivity' is possible; first, all scientific inquiry has purpose; it is pursued in a particular socio-economic context, within which certain theoretical or methodological positions are taken and particular problems are defined as worthy of investigation. Second, for research to be meaningful, it has to measure the properties of objects in order to be able to differentiate between them and subsequently test their relationship to one another. 'Objects' are not limited to that which is manifestly physical, for they also include phenomena such as institutions and mental states. Third and finally, the value of truth-seeking (the desire not to misrepresent, whether deliberately or by omission, the properties of the objects being studied) is shared across the sciences. Research may thus be both value-laden *and* objective; it is cognizant of the values that underpin its particular form of inquiry, yet does not simply produce knowledge that reflects these values.

1.5 Ontological positions - Idealism and realism

Thus far, I have traced a course in the philosophy of science that allows for the conduct of scientific research to be both contextualized (in a historical, social, and cultural sense) *and* rigorous, but have yet to consider ontology and epistemology. Heated debates about epistemology in the social sciences arise periodically, and on the whole are settled to the satisfaction of none of the combatants. The manner in which these debates produce more heat than light is arguably due, at least in part, to the absence of discussion surrounding the different ontological positions that the participants hold. Given that an ontology describes a belief about the way that the world must be in order for knowledge to be possible, it seems relatively clear that epistemology must be grounded in this ontology; debates about epistemology would thus seem unlikely to be resolved without a recognition of this grounding.

In view of the scant consideration that different ontological positions receive with respect to understanding why different research methods (as encapsulated in a particular epistemology) are advocated, two key ontological positions will be expanded

upon here; idealism and realism. In doing so, it is important to note that it is primarily the work of realist philosophers of science (particularly Roy Bhaskar and Andrew Sayer) that will be drawn upon as not only a means of exploring realist ontology, but of describing how idealism differs from it. This is done not because of a bias towards a realist ontology, but simply because the idealist ontology remains largely unelucidated by those who hold it. Indeed, the absence in idealism of an ontological account of how knowledge can be applicable outside of the time and space in which it was constructed is a key realist criticism (Bhaskar, 1978).

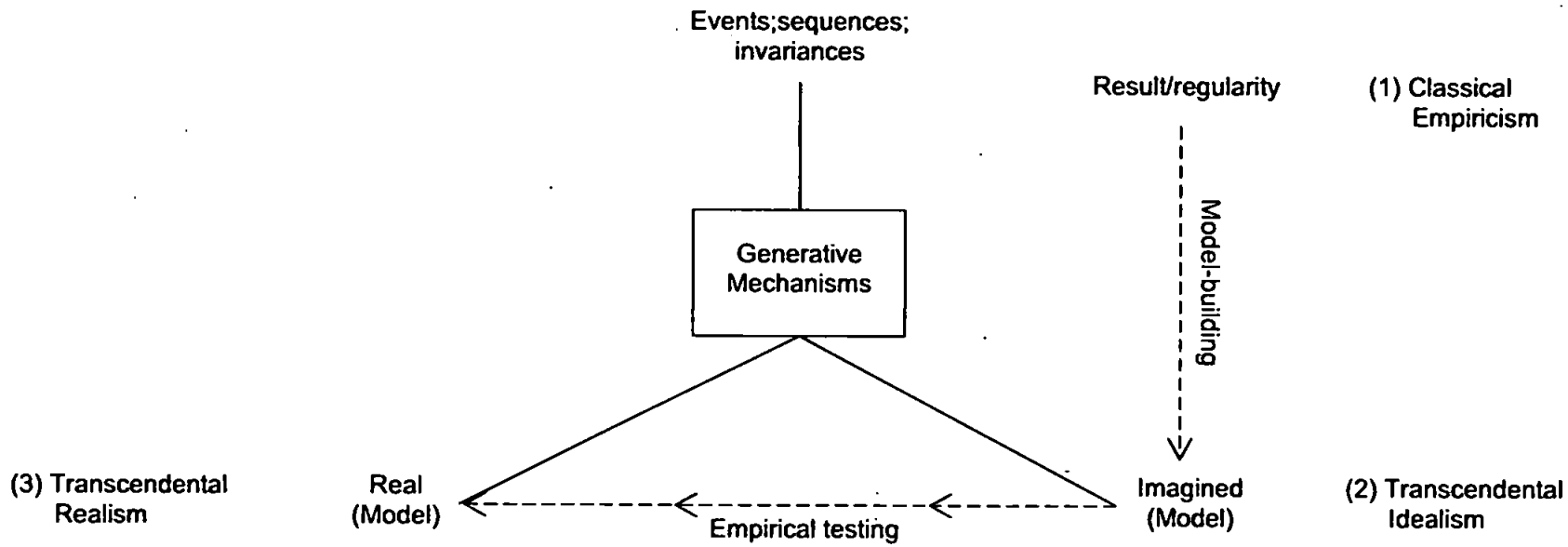
To understand realist ontology requires that it be placed within the context of the development of scientific inquiry as a whole. Bhaskar (1978, p.145-146) presents an outline (Figure 1) of how scientific inquiry has developed from a state of naïve empiricism (1) (the simple observation of regularities in phenomena) towards transcendental idealism (2) (where plausible explanations for constant conjunctions between phenomena are devised); Bhaskar furthermore makes the case for continuing the development of scientific inquiry towards transcendental realism[3] (3), where the mechanisms that underlay the relationship between phenomena are identified by empirical testing. This step beyond idealism is taken because, in realist ontology, these mechanisms are *real* and not simply plausible (but imagined) explanatory models. Even though such mechanisms are considered to be real, they are not observable in a straightforward sense; rather, their existence can only be discerned by means of careful testing to determine the contexts in which they do, or do not, operate (Bhaskar, 1978). The constant conjunction that provides the basis for a causal law in idealist ontology is considered inadequate in realist ontology, which believes that the world is too open a system to reasonably state that constant conjunctions will persist outside of the research context (Sayer, 1992, p.122-123).

[3] Transcendental realism is substantively different to empirical realism; the 'empirical' variant is limited to the *observable* properties of atomistic objects of knowledge. These objects are treated as possessing no deeper structures or powers (Sayer, 2000).

The progression beyond the search for constant conjunction between phenomena is argued to be evidence for the 'ontological depth' of realism. Rather than the analysis of outcomes that are observable at the surface, realism holds that the objects of knowledge with which one should be concerned are the (more deeply located) generative mechanisms that explain why events occur in the way that they do (Sayer, 2000). In contrast to idealism, which treats the world as consisting of patterns that exist between atomistic variables, realism treats the structure of the objects of knowledge as key in the formation of generative mechanisms, which themselves operate within a particular socio-political context (Sayer, 2000). This investigation of the deeper generative mechanisms is held to provide a greater understanding of the relationship between events, and moreover an ontological case for the continuing operation of these mechanisms (given particular contexts) outside of the experimental situation in which their operation was established (Bhaskar, 1978). As mechanisms are distinct from the pattern of events, the outcome of their operation can reasonably be different as a result of taking place in different contexts or where other mechanisms alter their operation (Bhaskar, 1978; Sayer, 1992). Importantly, mechanisms can contain powers that lie latent, meaning that the observation of a pattern of events (as in idealism) is considered insufficient for making the claim that they will continue to demonstrate the same pattern in the future; the operation of other mechanisms may stimulate powers within a key mechanism and thereby lead to a quite different outcome (Sayer, 2000).

As is strongly suggested by its name, a realist ontology lays claim to the existence of a real world about which it is possible to construct knowledge of the interplay of mechanisms within it (Bhaskar, 1978). In doing so, it lays the ground for a wide-ranging epistemology that is predicated upon the nature of the objects of knowledge to be investigated; in short, whilst the means of constructing valid knowledge are shared between the natural and social sciences to a certain degree, there are key areas where the nature of the social world differs so substantially

Figure 1: The logic of scientific discovery (Source: Bhaskar 1978)



that research methods should differ accordingly. For example, Sayer (1992) identifies the manner in which the social object of knowledge can learn and adjust to the researcher's interventions in a way that non-sentient objects of knowledge in the natural world simply cannot do. It is for this reason that reflexivity is considered important in realist investigation in order that the nature of this reactivity and its effect on the operation of mechanisms can be better understood (Sayer, 1992).

The comparison of realist and idealist ontologies has been made here in order to facilitate an understanding of why, with regard to social policy, a realist and an idealist would be likely to advocate different methods for evaluating the effectiveness of interventions. However, these portrayals should be read with the understanding that the day-to-day decision making that is rooted in these different ontologies can be substantially less clear cut than is suggested by the abstract discussion presented here. Nevertheless, these representations of the different ontologies can inform understanding of decisions made regarding how social policy interventions are to be evaluated (Pawson and Tilley, 1997, p.34).

What are the implications of a realist ontology for evaluating social policy, and for subsequentlycumulating this knowledge? The most complete development of a realist approach to evaluation in this field is that of Pawson and Tilley (1997); at its simplest, this approach advances an understanding of *outcomes* that follow from *mechanisms* which operate in certain *contexts*. This logic is grounded in a schema quite different to that of the idealist, where causal relationships are proposed to be best established by isolating variables in order that causation can be established (or not) with respect to the intervention. For the realist, it is the *mechanism* itself that requires testing, and this requires that experiments be designed in an effort to make the relationship between phenomena work in the manner in which they are proposed to (Pawson and Tilley, 1997, p.59-60). In this way, it is argued, mechanisms that operate outside of the confines of the experimental set-up (and hence have much wider

applicability in the field of social policy) may be discerned. To understand better why it is claimed that a realist approach could better contribute to the field of evaluation in social policy, the following paragraph details two key aspects of social interventions (embeddedness and mechanisms) about which a realist intervention offers particular strengths in terms of rigorous evaluation.

The nature of interventions in the social arena is that they are *embedded*. Objects and individuals operate within a social system that, as a whole, both limits and enables certain outcomes. Stated another way, these objects and individuals are not discrete entities between which cause and effect can be established, as the outcome is highly context-dependent. From the idealist's perspective, it may be argued that if different contexts produce different outcomes then interventions require testing in multiple contexts. However, for the realist, this is inadequate as a means of developing an understanding of how an intervention works. This is because contexts are not simply institutions or geographical locations, but sets of social rules, norms and values that impact substantively on the manner in which mechanisms operate and subsequently the success (or otherwise) of an intervention (Pawson and Tilley, 1997, p.64-65 and 69-71).

The manner in which *mechanisms* operate is also crucial to the realist process of evaluation. The importance that is attached to the discernment of these mechanisms is rooted in the ambition of realist explanation to establish accounts of generative causation that may be applied rigorously outside of the contexts in which they were produced. Rather than the idealist account where associative regularities are the means by which causation is established, in a realist evaluation it is the mechanisms themselves that are the regularity; in short, it is considered that regular association alone proves nothing, whilst the regular demonstration of the operation of a mechanism is substantive knowledge. Understanding these mechanisms is the essence of realist evaluation; it is contended that only by understanding the *way* in which social

interventions have different effects (or no effects) that the outcomes of these interventions can be made sense of and contribute to the cumulation of knowledge (Pawson and Tilley, 1997, p.65-69).

The following sections (2.0 and 3.0) consider the way in which knowledge cumulation, in the fields of health and social policy, has evolved in the UK. This juncture is a useful one at which to briefly summarize the way in which a realist approach to evaluation and knowledge cumulation is able to accommodate some of the key arguments made in this introductory chapter regarding the construction of knowledge. First, realism is cognizant of the important role played by theory in framing investigation of the objects of knowledge; rather than attempting to establish causal pathways between atomistic objects, the realist works within a framework that is mindful of the political and social context within which evaluation and knowledge cumulation takes place. Second, this acknowledgement of the inherently social nature of evaluation and knowledge cumulation means that realism is able to work pragmatically with inductive arguments, rather than seeking to justify deductive arguments. Finally, again recognizing the social nature of research, a realist approach does not seek to produce or synthesize wholly 'objective' knowledge, but is does allow for rigorous knowledge to be constructed and situated with due regard given to the role played by values in the process. This summary should not be interpreted as the simple advocacy of a realist ontology as being superior for the construction of evidence-bases, but it does set out why there are strong reasons for engaging with the realist approach to knowledge construction and synthesis in order to inform social policy and practice.

2.0 The evolution of Evidence-Based Policy and Practice in the UK

The contribution made by philosophers of science towards an understanding of the rigorous construction of knowledge, and the importance of ontology for understanding *why* different approaches to rigorous knowledge construction are

advocated, have been outlined in section 1.0. The following sections (2.1 and 2.2) trace the historical development and contemporary concerns of the fields of public health and the 'new public management'. An understanding of these fields is of substantive import to the analysis of knowledge construction in the case study SR contained in this thesis. It allows contemporary efforts to marshall evidence that might better inform policy to be placed within the context of efforts to do so over the past two centuries, and provides the basis for critically examining the contemporary claim that particular SR methods represent the most rigorous forms of knowledge production and synthesis. This critical analysis, with reference to EBM and EBP (for social policy) is presented in section 3.0.

2.1 The development of the field of public health

For the purposes of placing contemporary public health in context, the analysis presented here divides its development within the UK over the past two centuries into three eras. These eras cover the initial formation and medicalization of public health in the nineteenth century, the recognition of the roots of health inequalities in the 1970s, and the refocusing of policy in the late 1990s by the New Labour government and the subsequent promotion of the use of research evidence in policy making. In view of the importance of these more recent developments in the field, emphasis is placed upon the past thirty years. In tracing the history of public health in the UK, attention will be drawn to the degree that the focus of public health interventions, and the methods used to inform those interventions (namely, the evidence-base), have changed during that time.

From the time of its inception as 'public health' in Britain in the 1830s, conflict has arisen between professionals regarding control of the public health agenda. Edwin Chadwick, civil servant and author of the field's foundational text (which investigated the sanitary conditions amongst the poor of England) drew upon a wider European concern with the effects of urbanization, industrialization, and the free market upon the health of the poorer sections of society (Fee and Porter, 1992). Whilst Chadwick firmly

believed that technological advances (such as in the engineering of drains) should be deployed in the interests of producing the greatest health for the greatest number, he strongly argued that the implementation of such advances was not so much a technical as a political matter that required legislative change. By the middle of the nineteenth century, the rising power of a medical profession that was seeking new areas in which to make claims to expertise, resulted in public health reforms being somewhat restricted, largely to the development of an administrative structure that emphasized the role of 'state medicine' in the form of medically-qualified public health doctors (Fee and Porter, 1992). These doctors were predominantly involved in monitoring and reporting upon sanitary conditions rather than pushing for wider structural reform for public health (Fee and Porter, 1992). It may be argued that this focus has persisted into the twentieth century, with the field of public health demonstrating considerable political timidity in not advocating wider reform that would improve health outcomes (Lewis, 1993).

The dominance of Medicine within public health was notably critiqued from within the profession itself by McKeown (1979). Situating his case within a wider critique of the Cartesian duality of medical practice, McKeown argued that Medicine's focus upon the body prevented proper attention being given to the substantial role that external phenomena play in determining health. Moreover, Medicine's claims to having dramatically improved health in the twentieth century through innovations such as immunization and antibiotics were misplaced; technological innovations outside of Medicine, in concert with social policy, had improved food supply and protection from environmental hazards to such a degree that the role played by medical interventions *per se*, whilst still important, was not of the magnitude commonly attributed to them[4]

[4] McKeown's argument should not be interpreted, in the latter stages of the twentieth century and early twenty-first century, as meaning that medicine has no significant role to play in public health. Indeed, once improvements in sanitation, nutrition and housing were made in the first half of the twentieth century, it was technological advances in medicine that contributed substantially to improved health (Bunker, 2001). A pragmatic view on this would simply be that in public health it is not necessary to choose *either* environmental changes *or* technological advances; it is a matter of judiciously balancing both in order to achieve the most equitable outcome.

(McKeown, 1979). The wider determinants of health began to receive greater attention at the national level in health policy documents (for example, Department of Health and Social Security, 1976) and the landmark Black Report (see Townsend et al., 1992), both of which stressed the negative relationship between a person's social class at birth and their life expectancy.

The Black Report (Townsend et al., 1992) was unequivocal in its analysis: widening inequalities in health were strongly related to social and economic phenomena, with the better off benefiting disproportionately from health services. Whilst the Report was careful not to attribute a causal relationship between (for example) low income and poor health, it was energetically argued that the strength of the correlations between social and economic factors (*inter alia*, unemployment, environment, education, and housing) and poor health were such that it should be a priority to investigate how these factors influenced health deleteriously and what could be done in order to address these inequalities. Crucially, the Report noted that these areas lay outside of the traditional influence of health policy within the National Health Service (NHS); in other words, it was a broader social policy that had the real potential to positively impact upon health. The evidence in support of this approach[5] was further strengthened in 1988 by Margaret Whitehead's follow-up to the Black Report, 'The Health Divide' (Townsend et al., 1992).

The scene was set for a re-analysis of the focus of public health. It was argued that policy and practice should take the wider determinants of health seriously, that nominal differences between professional health and social sectors should not stymie effective collaborations that could decrease health inequalities, and that the field of public health should not baulk at addressing structural issues that impacted upon health (Ashton and Seymour, 1988; Rose, 1992). This approach was broadly accepted,

[5] Whilst there was substantial and rigorous evidence to support policy that would address the wider determinants of health, this tended to take the form of social justice critiques of a Conservative government that in its turn employed welfare state critiques. Attempts to establish a sounder evidential basis for public health policy became lost in this polarized debate (Kelly, 2006a).

even within quite traditional medical circles; for example, a *Lancet* editorial explicitly supported a move away from research that focused upon the individual, arguing instead that it was often more important to understand how wider structures and processes led to ill-health (Lancet, 1994). The importance of utilizing a diverse body of evidence to inform policy in this area was highlighted, in particular the danger that a prioritization of experimental designs could lead to attributing simple causal processes to phenomena that were significantly more complex (Lancet, 1994; Dean and Hunter, 1996).

The election of New Labour in 1997 marked the start of a 'Third Way' in British politics, treading a path between Old Labour's economic intervention and Thatcherite Conservatism's radical free market policies. Notably, New Labour acknowledged and pledged to address the wider determinants of health (Kelly, 2006a) within a framework that stressed the need for departments to co-operate in the delivery of policy rather than maintaining a focus upon their own discrete areas (Jones, 2005). As envisaged in the plan for NHS reforms (Department of Health, 1997), the claimed advantages of the free market were to be harnessed throughout New Labour policy in a way that benefited the whole population rather than the limited sections of society which it was claimed that Conservative policy had benefited. The recommendations of the Acheson Report (Acheson, 1998) substantively informed New Labour's first key health policy document (Department of Health, 1999), which sought to establish an agenda for addressing the steepening gradient of health inequalities whilst still allowing for the role of individual agency in decision-making[6]. The Acheson Report had stressed both the lack of evidence for many interventions that were intended to reduce health inequalities, and the limitations of controlled trials for evaluating the impact of upstream interventions. To address the deficit in the use of evidence in the policy making

[6] Whilst there was little change in terms of topics (smoking, obesity, alcohol consumption, sexual health, and so on) from the Conservative government's prior public health policy (Department of Health, 1992), there was a substantive change in terms of focus; evidence that the most advantaged in society tended to benefit disproportionately from public health interventions led to policy that explicitly noted the need to establish effective interventions for the most disadvantaged in society.

process, the Health Development Agency (HDA) was to be established in order to map the evidence-base in public health, set standards, and disseminate these in order that practice could be based upon the best evidence (Department of Health, 2001).

The establishment of a formal body for producing knowledge in public health was an important departure. Whilst it was in part a response to wider cultural changes, amongst them the decreasing deference towards professionals' knowledge and the ever-increasing volume of knowledge produced through research (Pietroni et al., 2003), it arguably also signalled a broad consensus surrounding the association between poor health and a variety of determinants linked to material and/or social deprivation. Arguments about whether or not such a link existed were effectively over; what was important now was to identify what could be done to genuinely improve the lot of people at the sharp end of widening health inequalities (Wanless, 2004). Constructing this evidence-base was not going to be straightforward; little research specifically measured impact upon health inequalities (Milward et al., 2003), key constructs such as ethnicity, disability, and place were under-developed (Graham and Kelly, 2004), and research rarely considered how the intervention evaluated might be applied in different contexts (Killoran and Kelly, 2004). Research in the field still tended to focus upon the more easily measurable indicators, thereby failing to engage with the more important (and complex) social forces that drove health inequalities (Beaglehole et al., 2004). Finally, as in earlier times, the importance of drawing upon a broad spectrum of research methods was highlighted. Whilst the Wanless Report (Wanless, 2004), acknowledged the role that the review methodology developed by the National Institute for Clinical Excellence (NICE)[7] could play in informing a similar process of knowledge synthesis in public health, it also cautioned against relying too much upon randomized-

[7] The HDA merged with NICE in 2005 to create the National Institute for Health and Clinical Excellence (also known as NICE). In a similar manner to its health technology and clinical counterparts, the public health arm of the institution aims to evaluate both the effectiveness and cost-effectiveness of interventions. However, it also aims to identify effective interventions (or programmes) at the level of the environment as well as at the level of the individual (Kelly, 2005).

controlled trials (RCTs) as the sole means of establishing effectiveness when evaluating public health interventions.

Two themes recur in the history of public health in the UK. The first is the uncontroversial nature of the association between material and/or social deprivation and poor health. Whilst the precise causal mechanisms are rarely well-understood, few seriously question that the association is spurious. Against this backdrop, the continuing efforts of figures within the field of public health (both medical and non-medical) to orientate the focus of interventions towards the wider determinants of health seem justified but ineffective in the face of deeply-rooted social structures. It is debatable whether the New Labour government's manifesto commitments to addressing inequalities have been borne out by a genuine challenge to structural determinants of health, although it should be acknowledged that the ban upon smoking in public places (implemented in July 2007) represents a bold challenge to vested commercial interests.

The second theme relates to how evidence is constructed in order to inform practice in public health. The strong influence of the medical profession within public health has extended to the promotion of particular research methods (in particular, RCTs) that are argued to have been proven in clinical medicine as the least biased and most rigorous means of producing knowledge about the effectiveness of interventions. The importance of utilizing RCTs as part of a repertoire of research methods, and of not relying upon RCTs as the sole means of establishing effectiveness, has been made repeatedly in the recent history of public health[8] (McKeown, 1979; Lancet, 1994; Dean and Hunter, 1996; Acheson, 1998; Green and Tones, 1999; Whitehead et al., 2000; Davey Smith et al., 2001; Department of Health, 2001; Beaglehole et al., 2004; Wanless, 2004), yet the utilization of knowledge that has not been produced using

[8] This should not be taken to mean that the authors referred to here discount the utility of RCTs; what is being argued is that judicious use of RCTs should be made with due regard given to their strengths and limitations for informing public health policy.

RCTs continues to be hampered by doubts over rigour. This situation remains rather perplexing; there is consensus upon the association between deprivation and poor health and the need to better understand how upstream interventions can halt, and potentially even reverse the steepening gradient of health inequalities. However, the research methods that continue to be considered the most rigorous and unbiased for investigating these phenomena are simultaneously identified (and not just by individual researchers, but by government reviews such as those led by Acheson and Wanless) as only being suited to the evaluation of a limited set of public health interventions that focus upon the individual rather than the wider determinants of health. A swing to the other extreme (evaluating structural determinants of health alone) would be equally undesirable; a repertoire of research methods are arguably required to be utilized in the effort to understand the role played by both the agency of the individual and the structure of the society in health outcomes (Kelly, 2006b).

2.2 The 'New Public Management'

To understand why EBP in the UK developed as it did, it is necessary to understand the changes that the implementation of 'New Public Management' (NPM) techniques in public services resulted in and the wider political and economic environment which shaped those changes. It is important to do this in order to place EBP within its historical context, and to signal clearly how its genesis and proliferation were not simply the outcome of a nominally rational progression so much as the result of a confluence of factors.

The socio-economic malaise of the UK in the late 1970s provided fertile ground for Friedman's (2002) argument that free markets represented a superior means for expressing individuals' wants and needs than the ballot box (Green, 1987; Gray, 1993). This articulated with the strong anti-collectivist stance of the Austrian school of economic thought that had achieved consensus in the Conservative party (Bosanquet, 1983; Gamble, 1986) and resulted in an abrupt shift in the UK economy, from one

which adopted an interventionist Keynesian approach to one where a monetarist model predominated. Deregulation, privatization and reductions in taxes were aggressively pursued as a means of addressing populist discontent, where high inflation, unemployment, and a rise in trade union power were perceived to have impacted negatively on the economy (Bosanquet, 1983; Levitas, 1986; Gray, 1993; Gray, 1999; Bradshaw and Bradshaw, 2004). The post-war consensus surrounding the welfare state in a capitalist society was radically fractured; no longer was the provision of services for the well-being of the populace by government seen to be self-evident. At an organizational level, professionals were viewed as requiring active management rather than a reliance upon professional codes to regulate practice in the best interests of the public (Clarke and Newman, 1997).

The management of professionals was just one, albeit key, part of the larger tranche of revisions that came to be collectively known as the 'New Public Management' (Hood, 1991). Advanced as a means of fundamentally addressing the wasteful, slothful, and unresponsive behemoths that post-war governments had become without resorting to damaging cuts in public service provision (Osborne and Gaebler, 1992), NPM promised to achieve greater quality from fewer resources by delivering only the services that were effective, and to do so more efficiently (Pollitt, 1990). Implementation of this management style required a fundamental reorganization of public services (Hood, 1991). First, managers assumed decision-making roles that were previously the preserve of senior practitioners; they were to be 'free to manage', to have the authority to exercise their responsibilities in the best interests of the organization (Clarke and Newman, 1997). Second, rewards (whether in the form of pay to an employee or in resources allocated to a unit) were to be linked to outputs; it was results that counted, not the adherence to procedures for achieving those results (Osborne and Gaebler, 1992; Osborne and McLaughlin, 2002). Third, competition as a means of raising standards was to be promoted by the separation of purchasers and providers and competitive tendering for services (Osborne and Gaebler, 1992; Deakin

and Walsh, 1996). Lastly, the parsimonious use of resources was vigorously promoted in the effort to deliver more effective services for less cost (Osborne and McLaughlin, 2002).

It is perhaps not surprising, given the revolutionary nature of NPM and the significant implications that it had for professionals' autonomy, that it has been subject to substantial critiques. By no means do these critiques provide a watertight argument against NPM as a whole, but they do focus attention on the areas of concern regarding the implementation of NPM in the public sector. Critiques of NPM focus upon three major areas; its lack of coherency, its suitability for the management of public sector services that are both complex and inherently political, and the evidence-base for its effectiveness.

NPM has been critiqued for its fundamentally contradictory nature. It promises to be all things to all people, simultaneously empowering consumers through the market, increasing autonomy for local service providers, and strengthening the control which senior management and government can exercise over the delivery of services (Pollitt, 1990). It is argued that the impact of NPM is far more partial, producing both benefits and disbenefits to different actors according to the local context, in particular when the actual implementation of NPM practices differ substantively from place to place (Hood, 1995). This reflects the nature of the delivery of any programme; exact implementation according to a blueprint rarely occurs, and to claim that NPM was implemented in such a way and that it constituted a 'new global paradigm' was to promote the idea of a cohesive approach where there was none (Hood, 1995).

The possibility of expressing the outcomes of public service delivery in measurable targets, in order that monitoring and review of those services could be undertaken, has also been questioned (Pollitt, 1990). For Power (1997), such assessments impoverish the management process; 'objective knowledge' obtained

through auditing is utilized uncritically, whilst the process places the production of such knowledge outside of the realms of critique. Moreover, the constitutive effects of such knowledge whereby services are revised in an effort to meet the auditable targets which are set are rarely considered; Power (1997) contends that this lack of attention leads to service provision being incentivized in such a way that it can damage or destroy the strengths of that service. A similar theme is taken up by Broadbent and Laughlin (2002), who question the ability of auditing techniques to measure the more subtle effects of the skilled delivery of services by professionals and the ultimate effect upon social welfare that might result through not appreciating the impact of such skills. None of the critiques suggest that the work of public service professionals do not require evaluation, but it is the focus upon applying certain forms of measurement in NPM that are contended to produce a highly partial knowledge of complex areas of practice.

Whilst the NPM is ostensibly a set of technical management systems, the implementation of these is argued to be a fundamentally political process. In the UK, upon the election of a New Labour government that proposed a 'Third Way' between capitalism and socialism, social issues were 're-branded'. This was done in an effort to make the issues manageable by NPM systems in the context of a modernization project that emphasized citizen participation, access to employment, strong 'communities', and the UK's participation in a globalized economy (Clarke and Newman, 1997; Newman, 2000; 2002). The Third Way conceptualization of society required some fundamental redefinitions to be made; for example, if unemployment is the result of inadequate opportunities, then fostering the development of businesses and training opportunities is key to solving the issue. This would allow 'socially excluded' people to join the mainstream society of the employed through their contribution of work-effort and spending power to the economy. Positioned in this way, the 'socially excluded' become the objects of a managerial calculus in which services can be planned, implemented and evaluated for their outcomes on the number of

unemployed. The conceptualization of people as 'socially excluded' differs *fundamentally* from other conceptions of unemployment. For example, 'structural unemployment' conceives of unemployment being as a result of ebbs and flows in the global economy, an economy that is structured in such a way as to require a reserve of unemployed people to keep wage inflation down. It is very clear that the latter conceptualization would not fit into the NPM system.

The evidence-base for the effectiveness of NPM is argued to be weak, relying more for its rapid dissemination on the inspirational nature of seminal texts such as Osborne and Gaebler (1992) than the actual substance of its proposals (Hood, 1995; Clarke and Newman, 1997). The claim that NPM systems represent a wholesale change for the better from what was claimed to be an outmoded and ossified bureaucracy is debatable. Broadbent and Laughlin (2002) make a strong case for NPM having actually intensified key elements of bureaucracy (in particular, the hierarchical control of professionals through rules and procedures) rather than replacing them with systems that flatten these hierarchies and empower the actors in the system. But, as Pollitt (2003) points out, it is not really a matter of being 'for' or 'against' NPM; there are clear examples of NPM systems having markedly improved certain public services, and equally there are areas of public services where its implementation has produced negative results. The difficulty of attributing causality to the NPM systems, as opposed to any of the many other factors that can influence the effectiveness of public services for the better or worse, is an inherent problem in the field (Pollitt, 2003). A pragmatic response is not to dismiss NPM as a whole, but to judiciously apply it in those areas where its strengths are likely to offer an improvement in public services.

3.0 Evidence-Based Medicine and the development of systematic review methods for social policy

The growth of NPM systems in the UK, where transparency of decision-making and the demonstration of policy effectiveness were considered paramount, provided

fertile ground for the development of EBP. In the UK, this drive towards systematically reviewing and cumulating evidence to inform policy and practice decisions is arguably rooted in the programme of work known as 'Evidence-Based Medicine' (EBM). EBM sought to create an infrastructure that could synthesize rigorous research in order to provide a sounder evidential basis for medical practice, and thereby challenge the perpetuation of harmful practices in clinical medicine as well as identify interventions that were genuinely beneficial (Mulrow, 1987; Oxman and Guyatt, 1988; Evidence-Based Medicine Working Group, 1992; Sackett et al., 1997; Clarke and Chalmers, 1998). This approach to knowledge utilization resonated strongly across a range of departments concerned with social policy within a UK government that was keen to consolidate its pragmatic credentials (Cabinet Office, 1999; Department of Health, 1999; Home Office, 1999) and provided the methodological model that formed the basis of SRs for social policy.

EBM could not be said to have directed the development of SR methods in the field of social policy in a straightforward way, but it was arguably of significant importance as the immediate precursor of, and stimulus for, EBP. To better understand the nature of contemporary debates regarding SR methods it is important to understand two key facets of EBM; the promise that the explicit utilization of evidence held in the form of cumulated effect sizes from RCTs in evaluating the effectiveness of interventions, and the role of this evidence in challenging the opaque views of experts. First, evidence was argued to have the potential to reduce medical doctors' reliance upon the vague concepts of 'experience' and 'intuition' when making decisions about treatment. This was proposed to make decision-making, in consultation with those receiving medical care, substantially more explicit. Most importantly, this was contended to allow both patients and those lower down the professional medical hierarchy to question (using 'evidence') the decisions of senior doctors (Evidence-Based Medicine Working Group, 1992; Marshall, 1997; Sackett et al., 1997).

Second, it was argued that in order to 'build' the evidence-base that would inform the practice of medical doctors and the ability of their patients to question treatment decisions, an objective and transparent method of assimilating research that did not rely upon the subjective interpretation of a single reviewer writing a narrative review was required (Mulrow, 1987; Oxman and Guyatt, 1988; Chalmers and Altman, 1995). Systematic reviews, where the literature search terms, the databases searched, and the method of grading (and excluding) research papers are explicitly documented, were proposed as the best means of achieving this. The strongest forms of evidence for inclusion in a SR of the effectiveness of interventions are proposed to be appropriately conducted RCTs, both for their strengths in addressing performance, detection, attrition, and selection bias (Glasziou et al., 2001; Higgins and Green, 2005) and for the suitability of their results to being pooled in order to give a single estimate of effect size by means of the statistical technique of meta-analysis (Hedges and Olkin, 1985; Egger et al., 2001a). Such an approach does not inherently exclude the utilization of knowledge produced using other research methods, but the means of synthesizing these different forms of knowledge was not elucidated, with the net effect that the majority of methodological development took place with regard to the synthesis of knowledge form RCTs.

EBM was by no means wholeheartedly accepted by the medical profession, and an overview of the critiques is provided in the following section (3.1) in order to contextualize the later critiques of EBP in the field of social policy. EBM is commonly portrayed as assuming a linear relationship between the production of evidence and its uptake into policy and practice. Whilst such critiques are not entirely unjustified, section 3.2 outlines the manner in which advocates of EBM could be argued to have a more nuanced approach to the relationship between evidence and policy and practice. A reminder is also made in this section regarding the essence of EBM in order to question whether critiques of a broader 'evidence-based' society (section 3.3) are really clear about what they are contesting.

3.1 Critiques of Evidence-Based Medicine from within the medical profession

The fact that the EBM movement came from within the medical profession, rather than it being imposed from an external source, was no guarantee that it would be uncritically accepted. Critiques of EBM centred upon two factors; its potential to undermine the professional autonomy of medical doctors, and the significant difficulties that would be encountered in applying knowledge that was tenable at the population level to that of the individual.

Autonomous decision-making in professional practice is highly valued, not simply as a means of defining one's professional role but also as an intrinsic part of practising in a considered and defensible manner. Critics of EBM were concerned that the utilization of professional experience and reasoning would be not so much downplayed as drowned out by evidence from meta-analyses (Tannenbaum, 1993; Polychronis et al., 1996). Others were of the view that professional authority was explicitly classed by EBM as arbitrary and subjective, which risked substantively undervaluing the contribution of experience to the practice of medicine (Charlton, 1997) and of uncritically accepting the evidence from 'imperfect' research over the judgement that could be exercised with the benefit of extensive professional experience (Miles et al., 1998). Concerns were also voiced over the manner in which the use of 'evidence' from meta-analyses without the benefit of medical interpretation, could allow non-clinical parties in the health service to question medical practice (Miles et al., 2000). These critiques of EBM can be seen simply as the playing out of the art/science debate within medicine that periodically occurs, in this instance with the scientific aspect of EBM becoming emphasized and the defence of the artistic aspect of professional practice being defended (Pope, 2003). This analysis may well be correct, but nevertheless the debate usefully informs an understanding of the growth of EBP in the field of social policy.

Critics of EBM contended that to utilize the evidence, as produced by a meta-analysis, in making a decision about medical interventions for an individual, required more than the knowledge of a pooled effect size. In particular, the lack of contextual information surrounding the interventions that were being evaluated in a meta-analysis, were argued to inhibit decision making in clinical practice (Carr-Hill, 1995; Feinstein and Horwitz, 1997). Concerns were also expressed that supposedly simple answers would be provided to what were in reality complex clinical decisions, and to mechanistically follow such prescriptions would be untenable in professional practice (Carr-Hill, 1995). A related critique raised the issue of whether EBM was sufficiently critical, beyond assurances of rigorous screening for internal validity, of the knowledge produced using RCTs (Feinstein and Horwitz, 1997). As the EBM movement developed, the recognition of the importance of other aspects of medical knowledge grew (Pope, 2003). However, whether or not issues such as the integration of professional expertise with knowledge produced using RCTs (or indeed other research methods) were sufficiently debated and developed remains a moot point.

A further aspect of utilizing meta-analyses for clinical decision making concerned the basis for applying data obtained at the population level to that of the individual. Tannenbaum (1993) identified the need for medical doctors to exercise judgement in their utilization of knowledge from RCTs or meta-analyses, as it was untenable for these aggregate results to be utilized in any straightforward sense in deciding upon care for an individual patient. The lack of certain information in meta-analyses (in particular when making decisions regarding sub-groups who were not in any of the original trials) was also contended to inhibit reasoned decision making (Feinstein and Horwitz, 1997). That the means of utilizing meta-analyses for decision making at the individual level was not further elucidated by the EBM movement was "problematic" (Pope, 2003, p.273).

3.2 Evidence for policy and practice in Evidence-Based Medicine

The EBM movement remained somewhat perplexed at critiques of EBM and, not without justification, would point to the original formulation of EBM. This stated the need for the utilization of expertise in applying evidence (Evidence-Based Medicine Working Group, 1992), and which was subsequently explicitly summarized by one of the pioneers of EBM as:

> "... [the] conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients." (Sackett et al., 1997, p.1)

Sackett was clear that evidence from a meta-analysis could never replace professional expertise, but that it was vital for decisions to be made that were cognizant of it (Sackett et al., 1996). However, the question remained as to exactly how a RCT or meta-analysis could be utilized to inform decision making regarding an individual patient. The root of this dilemma was acknowledged by some members of the EBM movement as follows; the results of a rigorously conducted RCT or meta-analysis can provide data on outcomes that are less prone to bias and therefore nearer to the real situation, but they are unable to answer the question of exactly which individuals would benefit from a (medical) intervention (Davey Smith and Egger, 1998). It is therefore necessary for medical doctors to draw upon their professional experience in order to make inferences from the results of a RCT or meta-analysis as to whether or not a particular patient will be likely to benefit from an intervention.

The issue of how professional expertise is utilized within the medical profession in order to judiciously utilize evidence in decision making is an area that has only received a little attention from within the EBM movement itself. The emphasis has been upon providing justifications for the reasons why a meta-analysis beneficially minimizes the breadth of inferences that a practitioner has to make; in short, a meta-analysis is seen to answer the question of effectiveness at the population level, thereby leaving the practitioner to get on with the task of integrating that knowledge with the values of

the patient and with due regard to the availability of resources (Sackett et al., 1997; Muir Gray, 2001). The results of a meta-analysis are generally considered to be robust across different sub-groups; if results are consistent then there is greater reason to believe that they are transferable across a wide range of sub-groups (Egger et al., 2001a; Glasziou et al., 2001; McAlister, 2001; Higgins and Green, 2005). Moreover, the investigation of whether or not particular sub-groups benefit more or less from an intervention is considered to be a risky undertaking; this is because a major rationale for conducting a meta-analysis is that by combining effect sizes it is possible to attain a value nearer to the 'true' result. Performing a sub-group analysis, in this view, would substantially undo a lot of the good work that the meta-analysis has done by reintroducing a greater risk bias from confounding variables and the play of chance (Davey Smith and Egger, 2001).

The approach towards expertise in EBM has arguably been ambivalent. On the one hand, it is denigrated as subjective and inadequately informed, but on the other it is considered necessary for the reasoned utilization of evidence produced by the EBM infrastructure. This points to the identification in EBM of a 'research-practice gap' that is to be filled by the provision of rigorous evidence. It has largely been assumed in EBM that knowledge diffuses in a linear fashion (Dopson et al., 2003) and that the priority (once rigorous evidence has been produced) is to disseminate that evidence and address organizational barriers to its implementation (Haynes and Haines, 1998; Eccles et al., 2001; Glasziou and Haynes, 2005; Haynes, 2005a). This straightforward model of the diffusion of knowledge is largely considered to be inadequate outside of EBM (Nutley et al., 2002). Indeed, the utilization of 'new' knowledge has been shown in a variety of health-care settings to be significantly influenced by factors such as the ease with which it can be integrated with current tacit knowledge, and the professional lenses through which knowledge is interpreted and debated (Fitzgerald et al., 2005).

Whilst it is important to understand the critiques of EBM, and some of the limitations of its practice, it is arguably also important to be clear about what the intentions behind the movement were. In short, EBM (and by extension, EBP in other fields) was about providing practitioners and policy makers with more rigorous knowledge that could inform their decision making (Sackett et al., 1996). Although certain forms of knowledge regarding the evaluation of the effectiveness of interventions were undoubtedly prioritized, they were never intended to exclude other knowledge; moreover, it was considered to be vital to utilize these different forms of knowledge pragmatically to better answer practice and policy questions in the most rigorous way possible (Sackett and Wennberg, 1997). In considering the critiques of EBP in social policy, and the development of SR methods in the field, it will arguably serve us well to keep these points clearly in mind.

3.3 Evidence for social policy making

The utilization of evidence in social policy making, in the guise of EBP, proved contentious. There existed a widespread perception that methods for synthesizing evidence that might have been appropriate in medicine were being foisted upon a field that was substantially different. There was both a concern that social policy making would be required to draw upon only RCTs as a form of valid knowledge and that this evidence would not so much inform policy as actually direct it. In doing so, it was considered that large areas of important knowledge would be excluded from the social policy making process. In this section, the benefit of hindsight will be used to consider the degree to which these claims were tenable, and to look in greater detail at some of the development in SR methods that took place during the period in which EBP was being established in the UK (2002-2005).

Whilst it would be unreasonable to claim that key proponents of the EBP approach for social policy considered the policy making process to be a wholly linear affair, little attention was paid to precisely *how* evidence from SRs could inform policy

making (for example, see Chalmers, 2003; Oakley et al., 2003). In short, the view was taken that if more rigorous evidence could be provided then more informed policy could be made, even if that evidence could never be used directly to form policy. The real-world applicability of evidence was considered to be of great importance, for studies conducted under ideal conditions would not necessarily reflect the realities of everyday implementation of a process (Smith, 1996). In addition, at least in the field of health policy, there was a substantial body of work that acknowledged a range of influences upon the policy making process that extended some way beyond the straightforward consideration of 'evidence'. These wider influences contributed to decision-making that could be ignorant (either wilfully or by omission) of substantive evidence, and included the need for policy makers to draw upon their own experiential knowledge to judiciously balance competing demands, the role played by the composition of policy making committees and the processes by which these committees made decisions, and the sobering fact that evidence was often only able to offer partial answers to policy questions (Murphy et al., 1998; Black, 2001; Pagliari et al., 2001; Gough and Elbourne, 2002). The success of research in substantively influencing the policy making process relied less on the rigour of the original research (or SR) than the development of a relationship of trust between researcher and policy maker (Black, 2001).

Despite the more nuanced appreciations of the policy making process that existed, and which it was understood that EBP would contribute to the development of (rather than radically reform), there was widespread belief that the growth of EBP represented an unwarranted and dangerous revision of the policy making process. A technocratic approach to policy making which privileged certain forms of knowledge, ignored the vital interpretive aspect of policy making, and focused on the technical resolution of issues that required complex political, social, and moral judgements was feared (Clarence, 2002; Parsons, 2002; Sanderson, 2002). This was evocatively expressed by Parsons (2002), who proposed that:

> "... [EBP] marks not so much a step forward as a step backwards: a return to the quest for a positivist yellow brick road leading to a promised policy dry ground..." (Parsons, 2002, p.45)

Given the acknowledgement by proponents of both EBM and EBP that the policy making process involved more than the straightforward, technical utilization of evidence (Muir Gray, 2001; Chalmers, 2003), the notion that what was being claimed for EBP was a wholesale leap into a clear decision making world seems somewhat far-fetched. It is perhaps better to adopt a longer-term perspective that views EBP as a swing *towards* instrumental policy making in the context of a long-running debate over the degree to which social policy making is to be informed by social science research (Young et al., 2002).

The processes that take place in a field such as social policy can vary widely, meaning that a focus upon a particular area risks an inadequate acknowledgement of processes occurring in other areas. Nevertheless, focusing upon the work of a particular institution within the field *can* be helpful as a means of increasing understanding of how EBP was implemented in one area of social policy and to question whether some of the fears of an instrumentalist approach to policy making were justified; this focus should not be taken to mean that the same processes were in place throughout social policy in the UK. The following section briefly summarizes the work of a key knowledge synthesis organization for public health in the UK, the HDA. Doing so allows some of the methodological development of SRs that took place, and the manner in which this was influenced by considerations of the policy making process, to be traced.

The development of SR methods by the HDA differed substantively from the ideas that existed about SRs providing knowledge for instrumentalist policy making. That there are issues of judgement in the SR process was not contested; the issue is how to minimize the elements of subjectivity and how to make the review process as transparent as possible (Swann et al., 2003). In setting out the methods for conducting

SRs for social policy, these methods can only ever be considered to be sound principles by which the review should be guided, rather than strict procedures that must be adhered to (Swann et al., 2003). In conducting a SR, the limitations of all of the forms of knowledge which are being drawn upon need to be recognized; notably, RCTs might justifiably be considered to produce the least biased knowledge about cause and effect relationships, but knowledge about the application of these findings in real-world contexts may be just as important (Kelly, 2004; Kelly et al., 2005). These limitations point to the nature of the knowledge that policy makers require; it needs to show how the practicalities of delivering a policy in different contexts may substantively impact upon its effectiveness, and the manner in which a policy may differentially affect various population groups needs to be elucidated if health inequalities are to be addressed (Health Development Agency, 2004). To attain these goals requires that different sources of knowledge be *integrated*, a comprehensive survey of the potential means of doing so being provided by Dixon-Woods et al (2004).

The position which I take here is that social policy making cannot afford to leave the policy making process to the vagaries of the political wind, or indeed to the opaque relationships that facilitate the uptake of certain research findings into policy. In essence, there was something inherently worthwhile about the goals of EBM and EBP; there needs to be an infrastructure for the synthesis and utilization of evidence from diverse forms of research. To propose this is not to simplistically assume that evidence can *direct* policy in an instrumental fashion, nor does it belittle the realities of policy making that needs to take into account the views and wishes of the electorate; what it does do is make a strong and clear claim that research evidence can and *should* contribute more substantially to a rational policy making process in order that social inequalities can be better addressed. To point out the manner in which researcher engagement in policy networks facilitates the utilization of research is to highlight an important aspect of policy making, but errs towards accepting an imperfect policy making process for *what it is* rather than *what it could be*. The essence of EBP is

arguably that if rigorous knowledge (in any form) pertinent to a subject of social policy exists, then it behoves a democratic society to endeavour to make proper use of that knowledge in the formation of social policy.

Conclusion

This chapter has drawn upon the philosophy of science and sociological analyses of both the development of social policy and the management of professional practice. In doing so, I have endeavoured to build the argument that the construction of knowledge can never be wholly objective, but that this apparent lack of objectivity should not overly concern us as long as we make concerted efforts to *understand* and *manage* it. This argument is rooted in the Popperian understanding of the impossibility of theory-neutral observation; theories are built upon past knowledge, and this knowledge is itself based upon inductive arguments that are founded upon theory-laden observation, and so on. Knowledge cannot spring forth untainted by the supposedly biased hand of human investigators. Neither can knowledge be cumulated in an entirely 'neutral' manner, even if there might be good reasons for advocating certain methods as more suitable and rigorous to a particular task of knowledge production and cumulation.

If it is indeed the case that there is no superior (that is to say, neutral) foundation for the production and cumulation of knowledge, how is one to judge the rigour of knowledge claims? Given that epistemology is grounded in ontology, and ontology reflects a particular understanding of the way in which the world works, there is a significant risk here that epistemological relativism becomes acceptable. It was for this reason that I presented a comparison of idealist and realist ontologies in some depth; this provided an example not just of how epistemological claims are grounded in ontology, but also of the way in which one may better understand others' epistemological claims when their ontological position is clarified. This is of some considerable significance for knowledge production and cumulation, for it shifts the

focus of debate away from competing for epistemological supremacy and on to the development of an understanding of the ontological basis of others' knowledge claims.

At this juncture, with the case for adopting an approach to knowledge production and cumulation that is fully cognizant of epistemology and ontology having been made, it is appropriate to consider how the political-economic environment can shape the knowledge that is produced. In the field of public health, there existed a broad consensus that health inequalities were strongly linked to wider economic and social phenomena. However, the lack of an infrastructure for cumulating knowledge about how to address these inequalities was of concern; this articulated with the growth of a NPM philosophy that stressed the need for measurable outcomes, the development of transparent methods for the evaluation of outcomes and the standards of practitioners' decision-making, and cost-effectiveness. A range of senior public health figures had cogently argued for the need to utilize a diverse range of research methods to investigate how health inequalities could be effectively addressed. However, the emphasis placed by NPM upon measurable outcomes articulated with the strong emphasis in EBM upon RCTs as the least biased, most rigorous method by which knowledge could be produced. This contributed significantly to the situation whereby RCTs became the preferred means of knowledge production with regard to the effectiveness of interventions, with a subsequent concentration of energy upon how knowledge from these RCTs could most rigorously be cumulated; the cumulation of other forms of knowledge was not entirely neglected, but the development of methods took place largely in the context of, and using the language of, RCTs.

The spread of EBM methods of knowledge cumulation to areas of social policy prompted substantial concern within the social sciences. Critiques of EBM tended not to engage with debates within medicine over issues such as how 'evidence' (in the form of SRs) could be integrated rigorously with expert knowledge. Fears about a future of technocratic policy making were regularly aired, but whilst it would be

simplistic to assume that more knowledge would by default lead to 'better' policy making, the issue of how knowledge might be more rigorously and usefully cumulated seemed to get lost in the argument and counter-argument that ensued over claims to knowledge. It is within this melee that this thesis is situated; how can methods for cumulating knowledge for social policy making be developed within a framework that is cognizant of political and social complexities, whilst at the same time allowing that knowledge really can make a substantive contribution to a rational and equitable policy making process?

Summary

Wholly 'objective' knowledge is a chimera, but an understanding of the role of ontology in knowledge creation is not. Appreciating the reasons why others' ontologies may differ from one's own is a crucial step *away from* internecine epistemological clashes and *towards* constructive engagement. Whilst it would be naïve to think that knowledge could ever straightforwardly direct policy making, it would be equally foolish to disparage efforts to develop methods to rigorously cumulate knowledge that could inform the policy making process.

Chapter 2

The Development of a Spectrum of Systematic Review Methods for Social Policy

Latterly, key figures in the field of Evidence-Based Policy and Practice (EBP) in the United Kingdom (UK) have made substantive efforts to further the integration of different forms of evidence and to re-evaluate the appropriateness of the hierarchy of evidence developed under the rubric of Evidence-Based Medicine[9] (EBM) (for example, see Oliver et al., 2005; Boaz et al., 2006; Littlejohns and Chalkidou, 2006). Whilst such efforts are to be applauded, they arguably provide more of a springboard for debate than an in-depth consideration of a number of substantive underlying issues. This chapter endeavours to take the next step by offering an analysis of these underlying issues by expanding upon the areas of agreement and disagreement between different approaches to conducting systematic reviews (SRs) in social policy. Given the importance of fostering the development of SR methods (Dixon-Woods et al., 2006a), and the conflict that can arise over knowledge cumulation (notable examples being Chalmers, 2005; Hammersley, 2005), it is intended that this analysis should enable substantive debate upon the issues concerned. Without the development of this kind of understanding, I argue that SR teams are likely to maintain their 'methodological preferences' because they are comfortable with them, rather than because they are necessarily the best tools for the job (Boaz et al., 2006). By expanding upon the commonalities that do exist, I hope to provide a firm basis for critically examining the often polarized methods proposed in the different approaches to SRs. This critical examination of the methods is intended to highlight the underlying rationale behind the different methods in order to better understand *why* these polarized approaches are strongly advocated.

[9] For example, see the Oxford Centre for Evidence-Based Medicine's 'levels of evidence' (http://www.cebm.net/levels_of_evidence.asp). *Ceteris paribus*, in establishing the effectiveness of an intervention, the trustworthiness of evidence follows a clear gradation from expert opinion (weak) to systematic reviews of randomized-controlled trials (strong).

As a basis for the consideration of the different approaches to conducting SRs, this paper takes a step further back in order to analyze the work of Donald Campbell. This body of work is taken to extend from Campbell's initial collaboration with Julian Stanley (Campbell and Stanley, 1963), through his later collaboration with Thomas Cook on quasi-experimental methods (Cook and Campbell, 1979), his sole author papers and book chapters throughout the 1970s and 1980s, and up to his final collaboration with William Shadish and Thomas Cook (published posthumously - Shadish et al., 2002). I shall argue that despite Campbell's name being quoted approvingly in much of the literature on conducting SRs, there has been only limited engagement with the breadth of his thought on issues surrounding research on the effectiveness of interventions in the social arena and on the social nature of all scientific endeavour. Given the difficulties encountered in conducting SRs in the field of social policy and practice, I argue that a genuine engagement with Campbell and colleagues' work is of great importance if these issues are to be adequately addressed.

1.0 The work of Donald Campbell – Lessons for contemporary systematic reviews in social policy?

Donald Campbell's work on evaluative research designs in the field of social policy, and wider issues of scientific practice, risks being all things to all people if only looked at in part. For those who favour randomized-controlled trials (RCTs), much will be found in Campbell and Stanley (1963) to bolster the case for this approach producing the most dependable form of knowledge, whilst those who view the social world as too complex for true experimental designs to be implemented will also find much to support their case in Cook and Campbell (1979) regarding rigorous quasi-experimental designs. Furthermore, qualitative researchers may be keen to use Campbell's (1978) thinking upon the qualitative basis upon which all quantitative knowledge is built when debating the relative certainties of different forms of knowledge. Even postmodernists may find something in Campbell's work to support

their case in his guarded support for constructivists' work on the social conduct of science (Campbell and Russo, 1999).

The above broad summary risks giving the impression that Campbell's work lacked coherence, that he latched onto trends in the research field rather than ploughed a consistent furrow. Another possible interpretation is that Campbell 'moved on', that he abandoned his earlier experimental approach in favour of one that leant heavily towards a social constructivist position. However, these interpretations of Campbell's work are not tenable. What marks out Campbell's work is arguably his willingness to traverse supposedly incommensurable approaches in a genuine effort to foster the production of pragmatic knowledge. Campbell's advocacy of the experimental approach never diminished as such, it simply became a little more refined and tempered with the knowledge that uncertainties in evaluations arose not only from the details of statistical inference, but also from the social processes inherent in the conduct of scientific research. Campbell's goal as a methodologist was to "... [define a] course between the extremes of inert scepticism and naïve credulity" (Campbell, 1978, p.185), and it is in this spirit that I argue that the entire body of his work should be approached, neither discounting Campbell's pursuit of rigour as unattainable nor accepting his advocacy of particular methods as the final word on the subject.

This section will now endeavour to engage with the breadth of Campbell and colleagues' work in order to assess how it may contribute to the contemporary conduct of SRs in social policy. It will be argued that a significant number of the current conflicts over the best methods for performing SRs in social policy have already been substantively addressed in the work of Campbell and colleagues. This is not to suggest that a straightforward solution sits waiting within his body of work, but the argument is made that engagement with it would facilitate the contemporary development of SR methods, and moreover would represent a more comprehensive treatment of the literature relevant to the rigorous conduct of SRs in the field of social policy.

1.1 Validity – internal, external, and construct

Throughout his work, Campbell strongly advocated the experimental approach (in the form of RCTs, where possible) to evaluating the effectiveness of interventions within the field of social policy because of the strength of the approach in discerning causal relationships between phenomena (Campbell and Russo, 1999). For example, the majority of the text of Campbell and Stanley (1963) is devoted to describing different experimental research designs and their various capabilities in addressing threats to internal validity. The importance attributed to the experimental approach remains in Campbell (1978), even though the bulk of that text is about 'qualitative knowing'. Similarly, in Cook and Campbell (1979), the focus on quasi-experimental research designs should not be seen as a dismissal of 'true' experimental designs, so much as a concerted effort to improve the practice of quasi-experimental methodologies. Campbell's advocacy of experimental approaches should thus be understood both in the context of his grasp of the appropriate utilization of other methodologies and of his increasing consideration of the importance of external validity, and the inherent tension that exists between assuring high internal validity and high external validity (Shadish et al., 2002).

In his later work, Campbell's advocacy of the experimental approach is a little more guarded, and he concedes that a rigorous utilization of the method is best suited to tightly-controlled environments, such as within prisons (Campbell and Russo, 1999). Campbell also identifies a tension between the assignment of people to experimental and control groups and his vision for the 'Experimenting Society', the philosophy of which was egalitarian and voluntaristic. Despite these reservations, Campbell continued to advocate the judicious use of the experimental approach, where it did not conflict with these ideals, for the evaluation of interventions in the social sphere (Campbell and Russo, 1999, p.24-26). In a similar spirit of egalitarianism, Campbell also opened up the possibility for the involvement of non-researchers in debating research findings, pointing out that those with first-hand experience of phenomena are

often in the strongest position to critique the rigour of a piece of research and the causal mechanisms which it purports to demonstrate (Campbell and Russo, 1999, p.37).

Campbell arguably attributed an increasing amount of importance to external validity in his later work, but it would be doing him a disservice to claim that his earlier work did not consider it at all. The difficulties of generalizing research findings to other populations or settings are touched upon (but never really resolved) in Campbell and Stanley (1963), where generalizations are seen to inherently involve assumptions about the mechanisms by which causal relationships operate and thus the extent to which these relationships may persist in other contexts. Campbell and Stanley (1963, p.33-34) propose that through the scientific process, these mechanisms may be further tested, and theories about them revised or refined, but there was no straightforward 'solution' to the degree of assumption in making generalizations. Fundamentally, what was necessary was a further testing out of theories about causal mechanisms in a range of contexts, and with the knowledge that the social world continued to change and develop apace.

Although recognizing the difficulties of making generalizations from research findings to a complex, dynamic social world, Campbell was certain that doing so was a vital task to perform if a cumulative approach to knowledge was to be taken. Whilst Campbell was strongly aware that the social sciences were unlikely to be able to develop laws that were robust over time and place (as in the physical sciences), he argued that it was imperative that the social sciences should endeavour to develop theories based upon research that to a certain extent could be generalized robustly over time and place (Campbell and Russo, 1999, p.195). Campbell was also clear that statistical inference, in and of itself, was insufficient as a basis for generalization, particularly where changes in phenomena over time produced as much substantive difference in results as differences attributable to other factors (Campbell and Russo,

1999, p.116-117). Instead, purposive sampling upon explicit conceptual criteria that endeavoured to test the generalizability claims made was proposed (Campbell and Russo, 1999, p.121).

A key development in Campbell and colleagues' thought upon rigorous generalization was the concept of construct validity, as something distinct from external validity (Shadish et al., 2002). Whilst external validity is concerned with the robustness of causal relationships in different contexts, construct validity focuses upon the inferences made from samples to the 'higher-order constructs' utilized in a field of study. For example, whilst it may be useful to know that the direction of the relationship between the provision of smoking cessation services and the number of people stopping smoking is robust over different segments of the population (external validity), it is the validity of the constructs about the phenomena concerned (such as people classified by social class) that allow or inhibit rigorous generalization to other instances. Whilst not strictly analogous, there are substantial similarities between constructs and 'middle-range theories' (Pawson, 2006) in that both concepts seek to understand phenomena in a way that allows the key elements to be judiciously applied to other instances. For example, research upon the experiences of people dependent upon government benefits for their livelihood may have strong potential (*if* the constructs are valid) to inform the understanding of issues affecting the long-term unemployed (Shadish et al., 2002, p.65). Similarly, particular approaches that guide interventions (in the form of middle-range 'programme theories') may be usefully applied in other areas (Pawson, 2006, p.74-78), although the uncritical application of these approaches could be misleading.

Both construct validity and middle-range theory are argued to be of substantial importance for the way that they allow inferences to be rigorously made within a theoretical framework. However, to do so requires the constructs that are utilized to be sufficiently refined for the task at hand. Shadish et al (2002, p.66-68) provide an

apparently banal example about trees that provides a central insight into the nature of constructs and the importance of identifying the *prototypical* features of phenomena. A householder can quite legitimately classify trees as deciduous or coniferous on the grounds that they would like a property where they have to do as little clearing of leaves in autumn as possible. However, a forester would use the constructs of angiosperm and gymnosperm, which differ from the deciduous/ coniferous constructs; this is because seed structure is of greater importance for forest management than seasonal patterns of leaf fall. It is crucial to note here that neither construct is 'right' or 'wrong'. Arguably, both constructs are correct in that they identify and classify on the basis of the most important prototypical features of the phenomena for the purposes concerned. The utilization of constructs in the social sciences may not be as straightforward or clear-cut as it is in tree management; there is frequently a lack of consensus over what the prototypical features of phenomena are (Shadish et al., 2002). However, this should not discourage us, for it is an inherent part of the development of construct validity that the constructs concerned are debated within a field that constantly seeks to re-evaluate the appropriateness of its conceptualizations (Shadish et al., 2002, p.66-72). The difficult question is whether or not researchers are prepared to seriously examine the dominant constructs that are utilized within their field.

1.2 The transparency of research methods

Campbell's methodological proposals always sought to attain as much transparency as was possible, but he was also quite explicit about the substantive tacit element, both of his own methods and those of others. This tacit element was first outlined by Campbell in his work upon the qualitative basis of all scientific knowledge, which he suggested went mostly unrecognized simply because of its ubiquity (Campbell, 1978). One of Campbell's concerns here was that a spurious precision could be apparent in research that was based upon unexamined qualitative knowledge, for example where the coding of interview responses is subsequently quantified for

analysis. The impact that this qualitative knowledge (potentially a mixture of 'common-sense', professional training, and personal experience and reflection) has upon the results of a study (wherever it may lie on, for example, an experimental-ethnographic continuum) is rarely considered, yet it may be of substantive import (Shadish et al., 2002, p.29). Campbell argued strongly that these underlying assumptions should be investigated in order to make experimental research more genuinely scientific, his proposal for a 'project anthropologist' (Campbell, 1978) being realized in the work of social constructivist researchers who examined the social process of science within laboratories (Knorr-Cetina, 1981; Latour and Woolgar, 1986). Perhaps surprisingly, Campbell wrote approvingly of the work of these researchers (Campbell, 1986, p.112-119) and the manner in which they brought forth the social processes by which laboratory researchers "... [imposed order] on a chaotic welter of inconsistent and inconclusive observations" (Campbell, 1986, p.118). Campbell was clear that such social processes did not invalidate the knowledge produced, but that overlooking the role that these social processes played *did* reduce scientific validity, as their role was not made clear (Campbell, 1986).

1.3 The role of the research community

Campbell argued that the conduct of rigorous science relied more upon a process of critical monitoring by the research community than the following of explicit procedures (Campbell and Russo, 1999). Campbell referred to this as 'competitive cross-validation', a social process that requires scientific claims to validity to be justified to a sceptical community, rather than taken on trust on the basis of a belief in a particular researcher's honesty or competence (Campbell, 1986).

The critical attentiveness which Campbell viewed as so vital for improving the validity of scientific claims worked through three different mechanisms (Campbell and Russo, 1999). First, in a community in which members respect and value each other's contributions, Campbell posited that the rigour of scientific conduct is improved by the

fear of humiliation if others' attempts to replicate an experiment fail because of weaknesses in the original design. Second, the nature of the community is that researchers are motivated to compete with one another in the sense of ensuring that high standards of conduct are maintained. Third, the critical nature of the community is such that it relies upon maintaining rigorous standards for its reputation, and thus is more likely to disclose poor practice rather than seek to cover it up. A key factor in all of these mechanisms is the focus upon the community, the "disputatious community of 'truth seekers'" as Campbell coined it (Campbell and Russo, 1999, p.9), in entering into debate. It was not a matter of procedure assuring knowledge of higher validity, rather it was a contingent process of informed debate that necessarily considered both internal and external validity and through mutual persuasion arrived at an agreed upon conclusion (Campbell, 1986; Campbell and Russo, 1999).

1.4 The importance of pattern identification

Campbell was clear that it was in the nature of knowledge that it cannot be lifted unproblematically out of the context in which it was produced; necessarily, the interpretation of data calls upon a qualitative knowing both of this context and of the manner in which the data can be made sense of within the field of knowledge as a whole ('pattern identification'; see Campbell (1978, p.191)). Whilst Campbell strongly argued the case in his conception of the 'Experimenting Society' that one should be willing and able to change one's views in the light of evidence (Campbell and Russo, 1999), he did not propose that the way to do this was to start from isolated data ("the very reverse of dependable building blocks" (Campbell, 1978, p.191)). Rather, an informed and careful interpretation of these data, in which the manner in which pattern identification is utilized is made as explicit as possible, is contended by Campbell to be a superior (if more difficult) way to 'work from the evidence'.

2.0 Areas of agreement in approaches to conducting systematic reviews

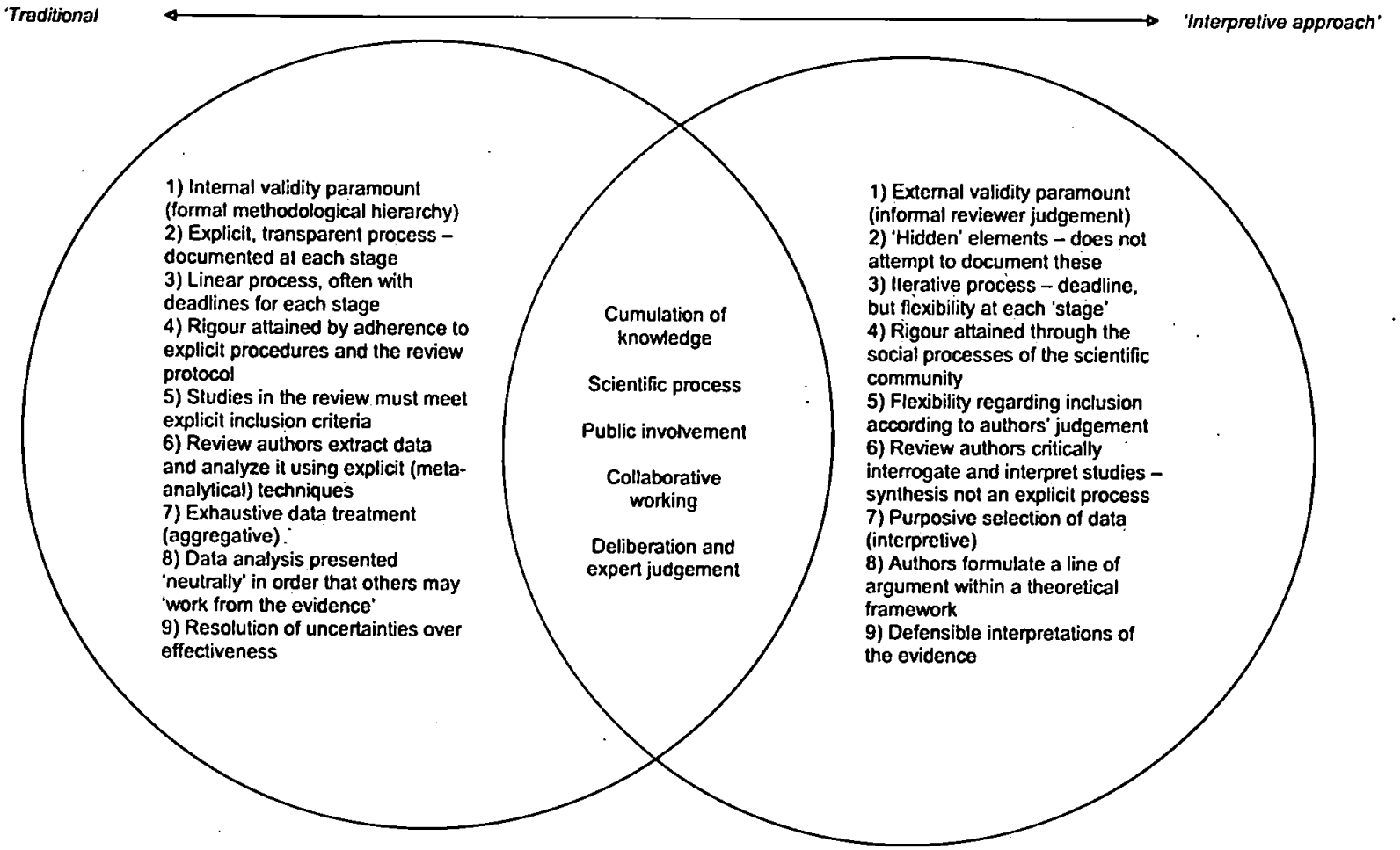
This section provides an overview of the current range of approaches to conducting SRs[10]. It is intended to be read with Campbell's consideration of the validity and transparency of research methods, and of the role of the research community, being borne in mind. First, the broad areas of agreement about objectives in the field, which are arguably greater than is generally acknowledged, are identified. By expanding upon these commonalities, I hope to provide a firm basis for critically examining the often polarized methods proposed in the different approaches to SRs. This critical examination of the methods is intended to highlight the underlying rationale behind the different methods in order to better understand *why* these polarized approaches are strongly advocated. The structure of this analysis, together with a visual representation of where some key approaches lie on the spectrum, is presented diagrammatically in Figure 2.

2.1 The cumulation of knowledge

A key aspect of the rationale for SRs is that policy and practice can be much better informed if research on a particular topic is analyzed comprehensively, rather than in a piecemeal and potentially partisan way (Chalmers and Altman, 1995; Sackett et al., 1997; Egger et al., 2001a). The social sciences have been criticized strongly for not developing an institutional structure that is supportive of such an approach, and the resulting lack of overall organization is suggested to result in the needless repetition of research, or of basing research upon knowledge that would have been discredited if a

[10] To date, a nomenclature for different approaches to conducting SRs has not been developed, and I have struggled here to find relevant terms. 'The dominant approach' may be suitable to a degree (novel processes are likely to be initially dominated by those who spearhead them), but the term is unhelpful for negotiating a productive path through the relevant issues. 'The EBM approach' is accurate insofar as it acknowledges the manner in which EBM pioneered the use of SRs in clinical medicine and how this approach was initially adopted as a blueprint for conducting SRs in social policy, but its use tends to suggest that EBM is a static phenomenon. I have elected to use 'the traditional approach' here to capture the idea that the first SR methods had certain features in common, but that later incarnations of these methods developed along a number of different routes.

Figure 2: The spectrum of approaches to conducting a systematic review



comprehensive SR had been undertaken (Chalmers, 2003; Oakley et al., 2005; Petticrew and Roberts, 2006). Moreover, this lack of institutional development is viewed by some as the result of an active resistance to methods which threaten established ways of working and vested interests (Oakley, 2006).

Thus far, there does not seem to be that much common ground between perspectives; the social sciences are viewed as having 'lost their way' with the advent of postmodernism, and of effectively throwing away a strong tradition of what is viewed as a genuinely scientific, cumulative approach to knowledge (Oakley et al., 2005). Whilst mindful of some of the more extreme postmodernist approaches, the more moderate social scientists that are the subjects of Oakley et al's critique do not appear, on further examination, to be so different with regard to the cumulation of knowledge. Writing in 1985, Martyn Hammersley voiced his concern that research in the sociology of education was proceeding along disorganized lines, with researchers failing to build upon earlier work, attempt to synthesize their own research findings with that of others, or work collaboratively (Hammersley, 1985).

Arguably, the differences between Oakley and Hammersley are with regard to the way in which they view knowledge cumulating. Oakley's position is shared with that of Iain Chalmers(Chalmers et al., 1989b)[11], where the computation of effect sizes literally cumulate; the more data one has, the narrower the confidence interval and thus the greater certitude with which statistical inferences can be made. Hammersley (2001) does not dispute this view of knowledge cumulation, but proposes that it is not the only way in which knowledge may cumulate – it can also be additive in the sense of forming a 'mosaic', or through challenging or re-enforcing other knowledge (Hammersley, 2001, p.548). Hammersley's position is thus shared with that of Freese (1980), who argued that statements regarding relationships between phenomena do not constitute a

[11] Sir Iain Chalmers has played a highly influential role in the EBM movement, publishing the first SRs in the field of obstetrics (Chalmers et al, 1989a; 1989b) and taking a leading role in establishing the Cochrane Collaboration, an international co-ordinating body for SRs of health research.

science, unless it can be demonstrated how these relationships operate. Such an understanding requires more of cumulation than is possible with a purely additive approach, but thedifficulties it poses for maintaining transparency and accuracy can be serious (Sharland and Taylor, 2006). Nevertheless, there is congruence between the approaches in that they both view it as essential that the production of knowledge takes place within an organized framework that prevents repetition and facilitates the utilization of prior research.

2.2 The scientific process

It could be argued that debates over SR methods are incommensurably rooted in different research paradigms and have moreover become mired in personalized exchanges rather than constructive debate (Chalmers, 2005; Hammersley, 2005; Oakley, 2006). Whilst such exchanges may indeed reflect the roots of the respective authors' research traditions (or, in the case of Ann Oakley, the development of her thought to outside of the notional borders of her research tradition), there arguably exists more common ground between the 'different' approaches than might be evident on first examination. This common ground exists in three areas; the underlying tenets of 'good' science, how theories should be tested, and the role of judgement in the scientific process.

First, there is common ground on how science should proceed; evidence should be provided for the conclusions that are reached, and methods should be reported so that the research community can critique the work reported upon (Oakley, 2000; Hammersley, 2001). Second, despite the apparent gulf between the experimental approach to research and, for example, ethnographic research, there is arguably more common ground than is normally acknowledged. Writing in 2003, Iain Chalmers summarized the rationale behind his long-standing advocacy of the experimental approach in both medical and social policy research. In doing so, he expanded upon the scientific approach whereby theory is refined or rejected on the basis of research

findings, and where the essence of the scientific approach with respect to establishing causation is the manipulation of variables so that relationships between phenomena may be observed (Chalmers, 2003, p.27). Almost twenty years previously, long before the fields of the two authors had come into contact, Martyn Hammersley was writing of the importance in educational research of the "... testing out of predictions derived from [a] theoretical idea" (Hammersley, 1985, p.247). Whilst Chalmers and Hammersley remain divided over the best means of establishing causation (for Chalmers, RCTs; for Hammersley, ethnographic research that tests out theory), the essence of manipulating phenomena so that the (conditional) effects can be observed or interpreted, remains.

Finally, there exists a nominal conflict over the role of judgement in the conduct of research, and in particular the conduct of SRs. Hammersley characterizes the 'positivist' approach as basing its validity upon the adherence to explicit procedures and rules, thus factoring out what Hammersley views as the inherent (and important) role of human judgement in the scientific process (Hammersley, 2001; 2005). However, Oakley asserts that such a characterization is unwarranted, as it fails to appreciate the "... messy, non-linear, creative [and] contingent" nature of the conduct of all science (Oakley, 2000, p.302). Whilst these two authors may still disagree over the extent to which judgement is actually exercised in scientific practice, there at least exists common ground between them regarding the role that human judgement, rather than uncritical rule-following, may play in the scientific process.

2.3 Public involvement in the systematic review process

Intricately tied up with the advocacy of SRs is the idea of democratizing knowledge. In a world where access to information in hugely facilitated by computer technology and telecommunications, and where there is an increasing public demand for the tools that allow devolved decision-making (as opposed to entrusting decisions to expert bodies), SRs have been promoted as a key method for synthesizing research evidence and making it available publicly. There is arguably broad agreement within

the field regarding the involvement of the public at various key stages of a SR, for example to refine the initial review question, to provide feedback on progress at interim stages of the review, and to comment upon the practical relevance of the final review's findings. Whilst the substance of efforts made to encourage and facilitate service users and the public to contribute to a review as equal partners (for example, see Braye and Preston-Shoot, 2005) are unlikely to be contested, a sociologically-rich understanding of the methods of consultation, how these affect the views elicited, and how these are integrated with expert knowledge is arguably missing. For example, Harden (2001) reflected upon some of the difficulties of eliciting and balancing views that were expressed in a review's consultation with young people, but further work on this problematic area has not been forthcoming. Nevertheless, the basic commitment to public involvement in SRs is shared across the field.

Despite this shared commitment to public involvement in SRs, it should be acknowledged that differences do exist with regard to accessing and utilizing the knowledge so produced. For example, Oakley et al (2005) view the collation of SRs in publicly accessible electronic libraries as essential for the democratization of knowledge. The view held is that to not do so simply protects the power of professionals, and places the service-user in a subservient position, unable to question the professional's 'expertise'. A different view is held by others; for example, Kuhlmann (2004) maintains that the democratization of knowledge needs to take place at a more fundamental level, namely at the stage at which what is acceptable as evidence or knowledge is decided upon. Whilst this substantive difference in the field persists, there is at least the potential for development in that both perspectives highly value the role of those outside of the 'expert community'.

2.4 Collaborative working

Criticisms regarding the lack of individual and institutional response to the call for conducting SRs have focused strongly on a perceived lack of understanding of the

rationale for doing so. Chalmers et al (2002) strongly criticize academic practice in the UK for not seriously considering the imperatives of research synthesis and how the challenges it presents may best be responded to. Oakley et al (2005) saw little development in academics' approach in the intervening years, criticizing academia in the UK for hindering the cumulation of knowledge by conducting research projects that were "parochial" and which lacked "... any sense of connectedness with one another" (Oakley et al., 2005, p.21). It is debatable whether these critiques apply as strongly now as they did at the time they were made; for example, the review approaches of Dixon-Woods et al (2006b) and Pawson (2006) clearly highlight the importance of situating a review in the context of other work in order that a piecemeal approach is not adopted. Pawson (2006) also goes a step further by specifying an approach in which collaborative working between reviewers and commissioners and policy makers enables the review to be kept highly relevant to the needs of these parties. Whilst the ideal of a wholly collaborative social policy review community has certainly not been attained, there is again the common ground that acknowledges the importance of collaboration upon which to build.

2.5 Deliberation and expert judgement

Systematic reviews, particularly in health care, have at times been unfairly portrayed as proceeding along purely rule-bound tracks that prohibit certain forms of knowledge from being exercised, such as that inherent in professional judgement (for an extreme example, see Holmes et al., 2006). However, the classical formulation of EBM explicitly acknowledged that decision-making should integrate more tacit expertise with the explicit research synthesis conducted in a SR (Sackett et al., 1996), and although the actual conduct of EBM *might* have emphasized the latter, key figureheads of EBM continue to highlight the importance of expert clinical judgement (for example, Haynes, 2005b) and of political judgement that is mindful of the role as representatives of the populace (for example, Muir Gray, 2005) in decision-making that is informed by evidence.

It is arguably now widely recognized that EBP has to make substantial use of both the syntheses of SRs *and* professional judgement in the deliberations that inform decision-making. However, a cautionary note should be sounded here, for although there may be a convergence in views over the importance of judgement in EBP, the stages at which it is considered appropriately exercised may differ quite markedly. For example, Chalmers' (2003) starting point is that judgements will have to be made that are mindful of resources and values, as well as "the evidence" (Chalmers, 2003, p.36), but judgements do not enter into the picture until the point at which 'the evidence' has been established. This sharply contrasts with Hammersley (2005), who contends that judgement is exercised constantly throughout the review process as well as in the primary research that is being synthesized. Hammersley foregrounds the role of judgement in order to highlight the manner in which exercising judgement is inherently problematic, but he also asserts that this does not invalidate it in the way that Chalmers views it to do. Thus, there are foundations upon which to build regarding the exercising of judgement in SRs, but also a number of substantive issues that require debate.

3.0 Areas of disagreement in approaches to conducting systematic reviews

The first section of this chapter endeavoured to seek common ground between some of the different approaches to SRs in an effort to provide a framework that would allow advocates of different approaches to see that their respective goals (if not their proposed methods) have more in common than is usually acknowledged. Where the approaches differed, I endeavoured to identify exactly where the disagreements were in a bid to foster substantive debate upon the precise issues in question. In this second section I do not attempt to reconcile different approaches to conducting SRs[12]. Instead, I adopt a different tack by clearly setting out the substantively different rationale that advocates of different approaches use in order to argue their case. Parts one and two

[12] It might be argued that the disagreements are more about epistemology than SR methods *per se*. However, building upon the more abstract discussion of ontology and epistemology in Chapter 1, my aim here is to focus the analysis upon the more concrete process of conducting a SR as a basis for the analysis contained in the case study.

of this chapter are summarized in Figure 2, where the polarized nature of approaches to SRs can also be seen on the continuum plotted between the traditional and interpretive approaches at the base of the figure. My motivation for demonstrating these polarized positions is not to contend that approaches to SRs are so inherently far apart that there is no chance of advocates of different approaches finding common ground; rather, it is an attempt to make clearer the rationale behind *why* different methods are advocated, in the hope that a better understanding of these will allow advocates of different approaches to engage with each other more constructively.

3.1 Internal versus external validity

> Internal validity:- "The validity of inferences about whether the relationship between two variables is causal." (Shadish et al., 2002, p.508)
>
> External validity:- "The validity of inferences about whether the causal relationship holds over variations in persons, settings, treatment variables, and measurement variables." (Shadish et al., 2002, p.507)

In view of the multiple sources of potential bias in the conduct of primary research, internal validity is argued to be the primary consideration in judging what research should be included in a SR (Chalmers et al., 2002). It is argued that bias will impinge on the process of research at a number of stages if steps are not taken to address it. Ensuring that research subjects are unaware of whether they are receiving an experimental or standard treatment (single-blinding), or both research subjects and investigators (double-blinding), is proposed to prevent performance bias (where subjects alter their behaviour in response to the treatment they are receiving) and detection bias (where investigators' assessment of results are biased towards their own views of what the results 'should' be) (Davies and Nutley, 2000; Egger et al., 2001b). Systematic differences between groups of research subjects may result in certain sub-groups being far more likely to drop out of the study than others; this attrition needs to be factored into the analysis if it is not to be skewed (Egger et al., 2001b). Finally, selection bias is overcome by the initial allocation of research subjects to intervention and control groups being randomized. This is supposed not only to prevent

investigators' assigning subjects to a particular group for unstated reasons (perhaps they may feel that a subject will benefit greatly from an intervention, and thus contribute to demonstrating an intervention's effectiveness), but also to produce experimental and control groups that are equivalent, with (assuming an adequate sample size) both known and unknown confounders distributed equally between the two groups (Oakley, 2000; Higgins and Green, 2005). It may also be argued that the greater the complexity of the phenomena of concern, the greater the need for random allocation to experimental and control groups in order to equally distribute these confounding factors.

The prioritization of internal validity has, however, been critiqued by those who are otherwise strong advocates of the role of randomization in attaining it. Glasziou et al (2004), whilst maintaining that randomization is the strongest method by which to attain internal validity, also clearly state that the ranking of studies on the basis of internal validity may lead to important qualities in other research designs being overlooked. These qualities may include a deeper understanding of people's motivations and rationale (qualitative research), the ability to evaluate an overall outcome over a long time period (cohort studies), and the identification of novel positive or negative effects that warrant further investigation (case reports).

More fundamental critiques of the utility of randomization have also been made regarding the logic behind the claim that randomization (if properly conducted) overcomes bias. First, the claim that evenly distributing confounders (both known and unknown) is adequately attained through randomization is contested by Alan Chalmers (1999). Chalmers suggests that the essence of the experimental approach is that every effort will be made to identify potential confounders, and an explicit account given of how those confounders are removed or controlled for. In this view, relying on randomization to distribute these confounders is an abrogation of the researcher's responsibility. Second, where research upon complex phenomena is conducted, there

may be important confounding factors which are *unknown*. If an indefinite number of these unknown confounders exist, how is a researcher to assess the probability of the experimental and control groups being unbalanced through the play of chance (Worrall, 2002, p.9-12)?[13]

Critique has also been made of the use of internal validity as the primary criterion for objectively assessing the rigour of research when conducting a SR. This critique considers the objectivity of this process to be illusory in view of the knowledge about the subject area that will be brought to bear on any assessment of research validity (Hammersley, 2001; 2004). In this view, objectivity is not attained through adherence to procedures for assessing internal validity, as to do so leads to an impoverished assessment. Utilizing judgement in the process is argued to enable a more informed appreciation of validity, but it is achieved by a process that is not explicit in the sense in which the Traditional approach requires.

Advocates of prioritizing the external validity of primary research focus upon the practicalities of everyday delivery of public services and the impact of contextual factors upon outcomes. Writing on the field of special education, Gallagher (2004) questions how a study using the experimental approach can establish that implementation fidelity was attained – how do researchers purport to have separated out the delivery of an intervention and, for example, the interpersonal skills of a teacher with a group of children? In asking this, Gallagher is making a case for the importance of external validity, and arguably for some less precise evidence that is more amenable to judicious application in other classrooms. Victora et al (2004) make a similar argument regarding evidence for public health interventions, advocating that evaluations of interventions in a variety of everyday settings accompanied by plausible

[13] Regarding this critique, see Byrne (2002, p.94); simplified, the law of large numbers states that, if a sample is large enough and adequately stratified, it is valid to consider the sampling distribution to be normal. *If* it is possible to draw such a sample, then Worrall's (2002) critique does not hold. Also, in a SR that utilized meta-analysis, the play of chance with respect to *unknown* confounding factors would have decreasing effect upon outcomes the greater the number of studies analyzed.

rationale for why the outcomes observed occurred, would better inform policy and practice than studies with the highest internal validity. Bambra (2005) highlights the difficulties inherent in attributing impact to specific policies (and thus their potential impact in other contexts) when the environment in which these policies are evaluated is already shaped by other policies which are themselves evolving; in a similar vein, Wallace et al (2006) point to the significant social and economic changes that can take place over time and which may confound external validity.

The importance of considering the effect of contextual factors upon outcomes is stressed by those who make the case for greater weight to be given to external validity. Hammersley (2005) makes the general point that in utilizing the findings of research, due consideration should be given to the representativeness of the sample of the wider population to which the findings are proposed to be applied. Such considerations of context are arguably rooted in work on realist evaluation, which posed the question, 'What works for whom in what circumstances?' (Pawson and Tilley, 1997; Pawson, 2002). This is especially important in the field of public health, where the focus upon internal validity has resulted in a lack of consideration of how key drivers of health inequalities (such as gender, ethnicity, or place) affect people's responses to interventions designed to promote health (Killoran and Kelly, 2004; Weightman et al., 2005; Asthana and Halliday, 2006).

Contemporary developments of methods for conducting SRs have pushed the role of external validity up the agenda. Oliver et al (2005) advocate a 'weight of evidence' approach where both internal and external validity are assessed for their adequacy in answering the review question and a judgement is made upon this basis of how much weight should be given to the findings in formulating the review's synthesis of evidence. Bonell et al (2006) propose that RCTs should routinely include process evaluations that highlight the degree to which findings may be generalized to a wider population, and which investigate how contextual factors may affect outcomes

and the mechanisms through which these effects occur. Whilst these approaches represent some convergence in thinking upon issues of internal and external validity, they arguably avoid consideration of more fundamental issues such as the appropriateness of the hierarchy of evidence and the precise nature of evidence synthesis.

More fundamental critiques and proposals for methodological development come from Dixon-Woods et al (2006b) and Pawson (2006). Both of these proposals endeavour to balance a (non-formalized) judgement of internal validity with consideration of how useful a particular piece of research is likely to be in answering the review question, and differ from conventional narrative reviews with respect to the efforts made to make the procedure transparent. These methods do not attempt the (logically implausible) use of research with both the highest internal validity *and* external validity[14]; rather, they allow for the judicious use of different parts of different studies by the reviewer(s) in order to develop concepts (Dixon-Woods et al., 2006b) or middle-range theories (Pawson, 2006).

3.2 Sequential versus iterative review procedures

A central concern of SRs is that they should bring the rigorous scientific process to bear upon the task of synthesizing evidence, and through doing so, eliminate the biases that have resulted in effective interventions remaining unrecognized, and ineffective (and possibly dangerous) interventions continuing unchecked (Chalmers and Altman, 1995; Egger et al., 2001a; Higgins and Green, 2005). The Traditional approach to SRs proposes that the appropriate way to do this is to have an explicit procedure formalized before commencing a review; this procedure should state the review's objectives and exactly how evidence will be sought, on what basis it will be

[14] Research inevitably involves tradeoffs; for example, random assignment can improve internal validity, but at the potential cost of failing to retain research participants who would increase the study's external validity (Shadish et al., 2002). The point being made is that the perfect piece of research is a chimera – what makes for good research is a pragmatic and judicious approach that weights internal, external and construct validity according to the particular demands for knowledge being made. See Shadish et al (2002, p.96-102) for a full discussion of this point.

included, and how it will be synthesized and analyzed (Chalmers, 2003). Decision-making during the review is suggested to be made explicit through the utilization of a data extraction form, upon which not only are data recorded, but also disagreements between reviewers and how these were resolved (Higgins and Green, 2005). This record may be utilized by others to inform a re-analysis of the data if the rigour of the review comes into question. These procedures are designed to best address the deleterious effects of bias, and, in conjunction with the play of chance being addressed by randomization, produce knowledge in which one may have more faith (Chalmers, 2003).

The transparency attained by adhering to an explicit, sequential process for conducting a SR has, however, been questioned on a number of fronts. First, with regard to meta-analyses, the sheer volume of data involved is noted to contribute to situations where there is a great risk of significant errors occurring, but which the critical reader has scant possibility of investigating and questioning (Slavin, 1995; Briggs, 2005). The volume of data is also argued to drown out the role that the reviewers' judgement played, again inhibiting critique of the meta-analysis conducted (Briggs, 2005)[15]. Second, and in relation to the Traditional approach as a whole rather than solely meta-analysis, the role of judgement on the part of reviewers throughout the conduct of the review is contended to be of significance. Hammersley (2001) describes how procedural adherence is likely to lead to distortions in analysis, as the critical capacities of reviewers (rooted in their wider knowledge about methods and the field in question) are of greater import. These critical capacities are not transparent in any straightforward sense, but are considered essential in a critical synthesis of research. Such a synthesis should, ideally, reflect a "... skilled and knowledgeable assessment of

[15] Formal procedures, as would be found in a meta-analysis, are advocated on the basis that they will assure greater rigour in dealing with the large quantities of data that are an inherent part of a SR. However, Briggs' (2005) critique is that the user of a meta-analysis has to largely take the results 'on trust' because of the manner in which (possibly significant) differences in study designs and contexts are subsumed in the wider analysis.

what is likely to be true" (Hammersley, 2005, p.92); it is not 'biased' for its 'lack' of explicit procedures.

The step-by-step, sequential progression of the Traditional approach is also contested on the grounds that insights may be gained in the course of conducting a review that should substantively inform the search strategy, the categories used in the analysis of data, and possibly even result in the revision of the original review question. The Traditional approach does not necessarily rule out this last development (for example, see Higgins & Green, 2005, section 4.6), but doing so is likely to be seen as indicating bias, meaning that it is an option which reviewers are reluctant to take. In contrast, review methods proposed by Dixon-Woods et al (2006b) and Pawson (2006) stress that although there is an overarching framework for a review that includes setting the review question, searching the literature and synthesizing and analysing the data, the process is *necessarily* an iterative one. Moreover, both Dixon-Woods et al (2006b) and Pawson (2006) emphasize the importance of reviewers' critical engagement with the research being reviewed, and of a going back-and-forth between the primary research and the (partial) syntheses thereof rather than a single round of data extraction.

3.3 Rigour: procedures versus the scientific community

The previous section referred to the importance attached to adhering to explicit procedures in traditional SRs. This section expands upon this rationale by looking at how proponents of different approaches to SRs contend that rigour can be assured. Traditional SR procedures were built upon a critique of the imperfections of the peer review process for publication[16]. Biases based upon extra-scientific factors, such as author or institutional prestige, the political leanings of a journal, and the desire to only publish studies showing positive results (Grayson, 2002; Godlee and Dickersin, 2003)

[16] It is seldom acknowledged that such critiques resonated with earlier work in the sociology of scientific knowledge. For example, see Mulkay (1991, p.44-46) regarding the 'extra-scientific' factors (such as time constraints and vested interests) that impacted upon the peer review process within a physics department.

were deemed to substantively skew what was published. Logically, the conduct of a SR could not proceed along such lines, and the utilization of explicit procedures in the conduct of reviews was contended to provide the safeguard against such biases.

Critiques of the use of explicit procedures in the conduct of SRs focus on two areas. First, Hammersley's (2005) prioritization of the role of judgement again focuses attention on the reflexive aspect of a reviewer's work in order to maintain rigour. Hammersley states that although adherence to procedure can improve the conduct of science, there is a point at which this adherence, if utilized without the requisite professional judgement, can become harmful. The role played by judgement also feeds into the second critique; what is the role of the scientific community in assuring rigour? In a traditional SR, this community is to be distrusted for its many extra-scientific biases. However, if a review is grounded in the interpretive tradition (as advocated in Noblit and Hare's (1988) meta-ethnographic approach), the scientific community's role is essential in providing a mechanism through which discourse on the topic in question, and the rigour of the research that informs it, can be pursued. This mechanism can also be utilized on a smaller scale within a review itself through the 'checks and balances' of a team of reviewers critically engaging one another during the process of a review (Dixon-Woods et al., 2006b).

3.4 Data extraction versus data interrogation and interpretation

In the Traditional approach, data is extracted from primary research in accordance with the procedure defined at the outset of the review. If the settings, populations, and interventions across the various pieces of primary research are sufficiently homogeneous, then a meta-analysis will be conducted. If significant heterogeneity is present, a narrative synthesis of the data will be conducted. Whichever route is taken, the Traditional approach draws a clear line between the objective presentation of an analysis (namely, the synthesis of research findings) and the

interpretation of that analysis (this being the task of, for example, policy makers) (Chalmers, 2003; Higgins and Green, 2005; Muir Gray, 2005).

Proposed SR methods that are grounded in an interpretive tradition have a substantively different approach to the utilization of data. Dixon-Woods et al's (2006b) 'critical interpretive synthesis' method shuns formal data extraction forms in view of their resource-intensive nature and the way in which, it is suggested, they inhibit flexible data extraction. Instead, a non-formalized record is utilized to inform a critical engagement with both the findings from the research and the theoretical frameworks that informed the conduct of the research. This may necessitate going back to papers to re-interpret them in the light of other work identified in the course of the review. Pawson similarly argues that the process of research synthesis and analysis is inherently interpretive and cannot be split into neat, demarcated phases where analysis and interpretation are separate; for Pawson, 'explanation-building' is the key activity in a SR, and the process of juxtaposing, reconciling, adjudicating, consolidating, and situating of 'middle-range theories' developed from the research (see Pawson, 2006, p.74-76) necessarily requires an ongoing engagement with the findings and the flexibility to re-analyze other research in the light of later research.

3.5 Aggregative data cumulation versus the interpretive development of new knowledge

The comprehensive treatment of all data from research that meets a review's inclusion criteria is a cornerstone of the Traditional approach. This comprehensiveness works in two ways. First, it is comprehensive in the sense that the review should be sensitive to the re-use of the same data (whether written-up in a different form by the same authors or cited in the work of others), as to compute effect sizes without factoring this in could effectively double-count (or more) the same studies (Egger et al., 2001b). Second, it is comprehensive in the sense that the review's search strategy should not simply include flagship publications, but also less widely-circulated (and non-English language) journals. The rationale here is that, because prestigious journals

like to publish positive research, results showing equivocal or even harmful outcomes tend to get buried in the literature, yet the evidence they contain may be of substantive import (Chalmers, 2003). The comprehensive treatment of relevant data, ideally (but not necessarily) in the form of a meta-analysis that combines the results of numerous studies (weighted according to validity if required) in order to compute a composite effect size, is thus intended to overcome the publication bias that is likely to occur if a purposive selection of data for analysis is conducted.

Proposals for SRs to take a more interpretive approach are based upon a quite different view of the way in which knowledge can cumulate in a useful form for policy and practice. An interpretive approach is more concerned with developing understanding of phenomena rather than calculating an intervention's effect size. As such, it claims to be more able to deal with contextual variation as it does not seek to remove the effects of context, but rather works *with* this variability (Dixon-Woods et al., 2006b; Pawson, 2006). These developments of the interpretive approach share a belief that much of the knowledge necessary for policy making and practice in complex social contexts is not aggregative in the sense of being able to sum its direction of effect. Instead, this knowledge is viewed as usable in a synthesized form only through a mental and social process that involves comparing, contrasting, and debating the concepts within the research and how they may or may not transfer into other areas of policy or practice.

3.6 Atheoretical demonstration of effectiveness versus knowledge synthesis within an explanatory framework

The final area for the comparative analysis of different approaches to SRs is with regard to what *sort* of conclusion a SR should have. Should it establish the (lack of) effectiveness of an intervention, or should it present a novel understanding of phenomena within a theoretical framework? If the position is taken that a (properly conducted) RCT "... provides more grounds than other approaches... for making causal inferences" (Oakley et al., 2003, p.171), and that such a study design, albeit

informed by a theory, is the strongest method for testing out the effect of an intervention (Chalmers, 2003), then it follows that the evidence produced assumes a level of truth of a higher order than that produced using any other method. For Oakley (2000), it is a matter of producing facts using the best methods available in order to address inequalities in society, for without these facts that provide evidence of inequalities there will be no change.

Other approaches to systematic review do not find this classification of theory (something devised based upon experience) and evidence (the demonstration of causality) acceptable. The rationale for both approaches is again tied up with views upon how amenable complex, open social systems are to experimental approaches. Whilst allowing a role for RCTs, a number of authors have foregrounded the importance of situating research syntheses within frameworks that facilitate the understanding of phenomena and the causal mechanisms that occur (Clegg, 2005). For Asthana and Halliday (2006), such a framework would be focused at an international level, where differences in, for example, political and social domains between countries would be utilized in an effort to gain an understanding of why a particular intervention may work in one country but not another. For Dixon-Woods et al (2006b), there should be a greater emphasis upon integrating research evidence into a coherent theoretical framework, and upon critically examining the adequacy of the frameworks that predominantly inform research, policy and practice in the area concerned. Lastly, Pawson (2006) similarly emphasizes the integration of evidence, in the form of middle-range theories, into a framework ('explanation-building'). Furthermore, Pawson develops a line of thinking that is considerably more complex in this framework than that allowed for in experimental research. Summarizing this is best left to Pawson's own pithy question: "*What* is it about this kind of intervention that works for *whom*, in what *circumstances*, in what *respects*, and *why*?" (Pawson, 2006, p.94, emphases in original).

Conclusion

The emphasis upon internal validity in the Traditional approach to SRs has arguably focused debate upon what is actually just one component of EBP. This focus has partly been a result of the manner in which research is reported in medical and health journals, but the adoption of the Traditional approach perpetuates the neglect of issues surrounding (for example) external validity (Glasgow et al., 2007). Continuing to focus solely upon internal validity will inhibit the substantive debate and methodological development that is required within the field of social policy (Dixon-Woods et al., 2006a). Whilst this chapter does not provide any answers as to whether certain SR approaches are superior to others with regard to social policy, it does set out the case for there being reasonable grounds for believing that a strong consensus about what constitutes a scientific approach to EBP can be developed. This requires a willingness to engage constructively with others who advocate approaches to SR that lie elsewhere on the spectrum (Figure 2).

To this end, Donald Campbell is arguably a key figure, although perhaps not in the way that he is conventionally presumed to be. Although there are many aspects of Campbell's work which could contribute substantively to the development of SR methods, two are of particular importance. First, Campbell's disposition provides a strong example for how the research community might better approach developments and debate in the field of SRs. Throughout his working life, Campbell maintained an open mind towards new and sometimes radical approaches to research, and he relished engaging in constructive debate over the strengths and weaknesses of these approaches. Second, Campbell clearly valued the critical role of the scientific community in maintaining standards, but also for its role in fostering the development (through critical debate) of research methods. This is a crucial aspect of the framework which Campbell's work provides for the development of SR methods because of the way that it placed value upon the judicious utilization of expertise; and whilst it was

entirely appropriate and justified to interrogate the basis of that expertise, this was unlikely to be attained simply by appealing to knowledge that was produced or synthesized by adhering to a methods protocol. This approach resonates with the Popperian understanding of the impossibility of theory-neutral observation, and the significance of acknowledging the way that epistemology is grounded in ontology[17], discussed in Chapter 1.

Third, Campbell and colleagues' work upon how research validity may be attained, and when prioritizing one form of validity over another may be justified, arguably provides the fundamental 'building blocks' for research (and subsequently, SRs) across the field of social policy. Thinking critically about whether knowledge in field of study is reasonably secure and consensually agreed upon, or whether it is the case that significant contention exists regarding that knowledge, can help inform the emphasis given in a SR to the different types of validity. From a wider perspective, the different forms of validity could act as very useful tools for investigating the ontological foundations of epistemological claims; at a fundamental level, this might involve a realist questioning an idealist's claim that the observation of a regularity is sufficient for imputing a mechanism to the relationship being posited. Whilst utilizing research validity as a means of investigating epistemological claims risks raising difficult questions, it should be noted that the concepts of internal, external, and construct validity are free of the usual 'paradigmatic baggage' that can enfeeble debate regarding SR methods.

Arguably, the development of SR methods *is* beginning to take account of the different forms of validity, although this has wavered upon the means by which a genuine synthesis of diverse forms of evidence can be attained. Syntheses concerned with more than the internal validity of studies have tended towards a framework of neat

[17] And most importantly, the way in which an understanding of how others' epistemological claims are grounded in a particular ontology can substantively contribute to debates that do not degenerate into internecine epistemological warfare.

complementarity between quantitative and qualitative research, where the quantitative synthesis provides the best estimate of effect size and the qualitative synthesis information regarding how best to implement the 'most effective' interventions in different contexts (for example, see Oakley et al., 2004; Brunton et al., 2005; Shepherd et al., 2006; Arai et al., 2007). Such approaches stop some way short of the more critical role that syntheses of qualitative research can play in subjecting studies judged to have high internal validity to critical scrutiny. Miller et al's (2007) synthesis of research upon women's experiences of breastfeeding is an exception, and a rare example of a SR that considers construct validity. The argument I am making here is that the critical and important role in knowledge cumulation that a review such as Miller et al's can provide would be substantively facilitated through a wider understanding, and utilization, of Campbell and colleagues' work on forms of research validity.

A greater engagement with the breadth of Donald Campbell's work by those working in the field of SRs could better foster debate regarding different approaches to SRs for answering social policy questions. Given the complexity of Campbell's thought and the dense nature of his writing, this is not an easy task. However, if the outcome is the capability to produce richer evidence syntheses that better shape practitioners' and policy makers' ability to make effective interventions in people's lives, then the effort will be justified.

Summary

There are considerably more areas of agreement between those holding different views on appropriate methods for conducting SRs than is commonly acknowledged, but there are also significant areas of disagreement. Making the effort to try and understand others' methodological rationale is likely to prove far more constructive than the advocacy of a particular epistemology. Donald Campbell's work on validity and the research process provides a framework within which constructive engagement over the development of SR methods can take place.

Chapter 3

UK Drug Policy Since the Late Nineteenth Century: Policy Networks and the Utilization of Evidence.

In order to understand how the systematic review (SR) analysed in this research was utilized in the making of recommendations for policy it is important to place this piece of policy making in its historical context. This chapter therefore seeks to provide an historical analysis of drug policy making in the United Kingdom (UK) that is set within its wider international context and which takes account of the role played by discourse in the policy making process. Berridge (2003) has identified the (not infrequent) situation whereby 'historical facts' become 'policy truths' via a process rooted in policy agendas rather than rigorous historical analyses (Berridge, 2003, p.518-519). In this way, the 'folk histories' of particular policy fields substantively frame the policy making process in a manner that is seldom acknowledged, and subsequently under-examined. For this reason, this chapter strives to attain an understanding of the process by which drug policy has been made; who were the main actors responsible and how did they come to a consensus upon what the problem was and how policy should address it?; how did the wider social and economic environment impinge upon the making of drug policy?; and how was 'evidence' utilized in the making of policy?

The chapter begins by considering the policy making process in the UK from a generic perspective before focusing in upon the rationale for an analysis of the development of UK drug policy based upon argumentation and discourse. This approach utilizes the framework proposed by Hajer (2003) to identify the story lines, myths, and metaphors that have been instrumental in the development of drug policy. In doing so, two key questions are raised; first, does an historical analysis such as this provide important evidence that should inform the contemporary drug policy making process?; and second, what are the implications of this analysis for the way in which

the phenomenon of substance misuse is framed and the nature of the evidence which is sought to address the problem as it is defined?

The analysis of the development of UK drug policy presented here is hinged around the concept of addiction. Of utmost importance and interest analytically are the ways in which the phenomenon of substance use has changed over time in response to a wide range of influences that are at times only tangentially relevant to, or even arguably irrelevant to, managing the problem of substance misuse. In summary, whilst addictions were not recognized as such in the early nineteenth century, a concept of *addiction as a disease* was developed in the latter half of this period in which its regulation was viewed as a largely unproblematic affair. It is only in the twentieth century that addiction (to substances) was transformed into a dangerous social phenomenon as use spread (at different times) outside of the confines of medical supervision into the non-dominant cultures of young women, minority ethnic groups, and the working class. If, as is argued by Levine and Reinarman (1991), US drug policy has based its penal approach on tenuous evidence at the expense of a more sober assessment of the risks of substance use, could it be the case that UK drug policy might also benefit from a re-analysis of how the 'problem' is conceptualized?

In order to make clear the basis for the analysis presented here, it is necessary to understand the approach taken towards substance misuse. The approach is informed by Becker's groundbreaking study of cannabis use (Becker, 1963). In this study, Becker argued that cannabis smokers had to *learn* how to appreciate the physical and psychological effects of THC[18] in order to develop habitual use. What may be experienced as nausea by the novice may be experienced as a desirable sense of detachment and relaxation by a habitual user. Becker also identified the importance of the social setting in terms of fostering the perceived effects of substance use. Such an analysis points strongly towards the need for addiction to be analysed in biological,

[18] THC (Tetrahydrocannabinol) is the main psychoactive ingredient of cannabis.

psychological, sociological, and historical terms if an adequate understanding of it is to be developed (Zinberg, 1984). As such, one needs to be clear in differentiating craving, tolerance, and withdrawal as (physical) symptoms of addiction from the accompanying widespread concept of addiction. This concept, which attributes particular substances with the power to destroy an individual's moral fibre (and subsequently the fabric of society)[19] tightly focuses attention on the substance itself, to the neglect of analysis of the social matrix within which substance use arguably needs to be understood (Hugh-Jones, 1995).

It is because of the multiplicity of disparate forces that have, over time, fed into the formation and promulgation of the concept of addiction that an historical analysis is presented here. If we can better understand how the concept of addiction developed, we might better be able to understand the lens through which drug policy is dominantly conceptualized and how drug users are positioned within that conceptualization. Using a term of Donald Campbell's, this chapter asks whether *construct validity* is attained in the dominant approach to substance misuse in the UK. If the intrinsic dependence-creating properties of substances are only *part* of the reason why an individual habitually uses them, there remain substantive areas of knowledge about the phenomenon of substance use to understand. If this argument is at least partially correct, it follows that the management of substance use needs a wider conceptual focus than simply endeavouring to minimize the numbers of 'at-risk' individuals taking these substances.

1.0 The policy making process

Before analysing UK drug policy in depth, it is important to consider more generic analyses of the policy making process. In this way, the making of drug policy can be positioned within an understanding of the wider public policy process; this understanding necessarily includes the relationship between research and policy and

[19] Although it should be noted that the concept of addiction was not the first to link substance use with moral breakdown. This point will be developed in section 2.1.

the means by which consensus is sought in the formation of policy. The insights into this process which may be attained using discourse analysis will also be considered, as will the suitability of such an approach when dealing with historical sources. It will be argued that although the nature of the sources regarding drug policy preclude using a purely discourse analytic approach, there are key elements of this approach that may be utilized to good effect in attempting to gain a fuller understanding of the process of drug policy making in the UK.

1.1 Perspectives on policy making

Before considering the different approaches that exist for the analysis of the policy making process, it is important to broadly outline the post-war historical context of the UK and the inter-relationship this had with policy making. This historical contextualization is vital for attaining an understanding of policy networks and for clarifying the role that scientific evidence has played in the policy making process in the UK. Moreover, the historical context substantively informs the analysis of contemporary policy in the era of Evidence-Based Policy and Practice (EBP).

The UK in the post-war era, until the election of Margaret Thatcher's Conservative government in 1979, was characterized by a broad political consensus. During this period, a change in government seldom resulted in substantial policy change; instead, policy was notable for its stability over time and the incremental fashion in which changes were made (Dorey, 2005a). This stability extended to the individual policy networks. Civil servants and actors outside of government generally shared the same view about the purpose and priorities of policy making in their particular field and enjoyed a symbiotic relationship. In this way, non-government actors provided expertise in exchange for the opportunity to influence the direction of policy, and civil servants could depend on non-government actors to co-operate in policy implementation because of the role they had played in shaping policy (Dorey, 2005a). The deep-rooted nature of this relationship effectively precluded anything other

than incremental policy change, as the direction advocated by an individual minister was firmly tied to the wider policy network and interests outside of the network were easily dismissed. Change, however, did occur quite dramatically in 1979 with the election of a strong and confident Conservative government that wasted no time in "confronting, destabilizing or marginalizing" those interests in policy networks with which it took issue (Dorey, 2005a, p.216).

The 1980s were thus notable for the strong ideological lead that policy makers received from government, although there were clear exceptions in certain areas (for example, with regard to AIDS/HIV policy, see Berridge (1996)). However, the wider social processes that the Conservative government helped set in motion also began to reshape the nature of government itself (in the UK, as elsewhere in the world) towards a model of *governance*, where self-organizing networks are managed in the process of making policy, rather than the more directive and controlling nature of *government* (Rhodes, 1997). This reflected the changing nature of economic and political loci of control which were moving away from the governments of nation states in both directions; trans-national companies and bodies (such as the European Union and the United Nations) were gaining power at a macro-level, whilst government departments (re-branded as 'agencies') and the devolution of power to the Scottish, Welsh and Northern Irish assemblies began to take away power at a more micro-level (Hill, 2000; Dorey, 2005b). Whilst policy making in the UK has arguably long been characterized by the pursuit of consensus rather than the straight imposition of power (Jordan and Richardson, 1987), these changes have resulted in a policy making environment that is significantly more open to the representation of a wider range of interests than previously, and one which places significant emphasis upon the co-ordination of different policy sub-systems without actually being centrally directed (Dorey, 2005a; b). The conduct of government is thus characterized by the management of the policy making process, itself reflecting a tension between retaining control and letting go. Whilst policy making has never been an entirely straightforward affair, the international

political and economic developments of the latter quarter of the twentieth century have resulted in the policy making process becoming considerably more 'messy' than previously (Dorey, 2005a).

Analysis of this 'messy' picture of policy making in the UK can broadly take two different forms. The first sees policy making as a rational, progressive process. Analysis in this approach is rooted in Simon's (1957) treatment of policy making as a process in which alternatives are logically considered for their strengths and weaknesses in meeting organizational goals. For example, Hogwood and Gunn (1984) delineate the sequential stages in which it is argued that policy making occurs; problem(s) are selected, sources of solutions to these problems are identified, the means by which these solutions may best be put into practice are decided upon, and once implemented, the outcomes of the policy are evaluated.

The rationalist models, inspired by Simon's (1957) work, in which scientific research provides the 'facts' upon which decisions were made, have been argued to be over-simplified and insufficient for explaining the complexities of policy making processes that are shaped by wider social, political, and economic phenomena (Smart, 1984; Berridge and Thom, 1996). Critiques of the rationalist model of policy making are not just a feature of recent academic work. For example, Lindblom (1959; 1979) contended that the process of policy making is largely incremental, reflecting a politically astute balancing of different interests and a pragmatic lack of certainty about the degree to which 'solutions' can be provided for complex social problems.

Critics of rationalist models argue that the policy making process does not proceed along a purely linear and 'rational' path that is directed by the scientific evidence; rather, contemporary political concerns, cultural values, and the policy history of the phenomena concerned are an inherent part of the process (Green and Thorogood, 1998). This does not mean that scientific evidence plays no part in policy

making, but what it does mean is that the manner in which this evidence percolates into actual policy is far more dependent upon the shared understandings of problems and proposed solutions that develop within networks of practitioners, policy makers and other groups[20] (Weiss, 1979; 1987; Berridge and Thom, 1996). The manner in which policy makers interpret the evidence, scientific or otherwise, is not a straightforward process in which the 'best' evidence directs policy; policy makers' understanding of issues are substantively informed by conversations with colleagues, the mass media, and prevalent social science theories, but also their perception of the degree of consensus that exists over an issue[21] (Weiss, 1979; 1987). Thom (2005) provides the apt metaphor of policy making as a game of football; evidence is the football itself (the game could not take place without it), but it is the inter-relationship between the players (policy makers) and the teams (professional or policy networks) that is of crucial importance.

In the course of the discussions that policy makers hold, it is necessary for the problems which they are discussing to have been defined and agreed upon. This necessitates the 'framing' of phenomena in certain ways that resonate with the wider policy environment. These frames are created out of the political discourse (itself reflecting the broader social history of the culture concerned) of a community or nation, and analysis of them is argued to be crucial for gaining an understanding of the manner in which they define what counts as a problem and how this subsequently shapes and limits the direction of appropriate policy responses (Howlett and Ramesh, 2003; Dorey, 2005b). Thom (2005) illustrates the effect of framing using the example of alcohol policy, making the case that contemporary 'alcohol in moderation' policy is strongly reminiscent of the long-ignored (since the late nineteenth century) temperance

[20] The policy network approach has been criticized for its failure to contribute to theory development in the field of policy studies, but it is acknowledged that as a metaphor for increasing understanding of instances of policy making it has utility (Dowding, 1995). It is in this heuristic sense that the concept is utilized here.
[21] In a somewhat confusing manner, Weiss terms the diverse (and often informal) ways that policy makers are informed about a particular topic as 'enlightenment' (Weiss, 1979; 1987), but then continues by stressing that these sources of evidence may be "... partial, over-simplified, inadequate, or wrong" (Weiss, 1979, p.430). This ambivalence notwithstanding, Weiss' work remains an important elucidation of the role played by networks in incrementalist, non-linear policy making.

movement's advocacy of moderate alcohol consumption. Thom argues that the wider framing of the 'alcohol problem' (as a moral failing at the turn of the twentieth century and a disease model in the 1950s) meant that the temperance movement's presentation of evidence for policy making fell outside of the 'frame' and hence was not considered. It was not until the movement towards a population-level approach, using the language of public health and conceptualized through, for example, consumption-harm theories, that the evidence for the effectiveness of a moderate consumption approach to alcohol was able to be comprehended within the policy making 'frame'.

1.2 Argumentation and discourse in the policy making process

The conception of policy making as an interpretive process outlined in the previous section necessarily requires that the role played by language is better defined. It is clear that an interpretive approach does *not* conceive of the use of language by policy makers in the course of their deliberations as a means of transparently representing the facts that have been established using the requisite tools (Throgmorton, 1993). Rather, language frames problems in particular ways and is the means by which *arguments* are constructed and pursued in the course of making policy (Majone, 1989; Fischer and Forester, 1993). If a fuller understanding of the policy making process is to be attained, then it is imperative that the construction and deployment of these arguments be analysed.

Argumentation can be defined as the persuading of others through the skilful use of language to marshal evidence and ideas that resonate with the concerns of the listener. The term often has negative connotations, being associated more with persuading someone against their better judgement rather than a genuine process of exchange from which both parties may learn (Majone, 1989). Moreover, the notion that the policy making process inherently involves argument and persuasion, rather than a clean split between factual knowledge (which is utilized) and values, judgements, and opinions (which are not), substantively challenges the instrumental rationality

conception of the policy making process (Majone, 1989). Indeed, to force this fact-value distinction onto the process is to impoverish understanding of the work that policy makers do. This work inherently involves both the analysis of evidence (broadly defined) and the articulation of arguments about that evidence (Fischer and Forester, 1993; Throgmorton, 1993).

To better understand the utilization of argument and persuasion in policy making it is necessary to analyse the manner in which language is used to perform various functions. Typically, it will delineate the mechanisms by which the processes under discussion take place, define how the constituent elements of these processes can (or cannot) be manipulated, and define who and what is significant in these processes (Throgmorton, 1993). The arguments constructed do not consist simply of evidence from past research or expert opinion; rather, they are a complex admixture of these in addition to mathematical and logical arguments, value judgements, and recommendations (Majone, 1989). Unless the complexity of the construction of arguments by policy makers is considered, and this construction is analysed as a craft rather than an exercise in formal logic, the understanding of the policy making process is likely to remain impoverished (Majone, 1989; Throgmorton, 1993).

It thus remains to make the case for the analysis of discourse as a key means of improving understanding of the policy making process. Hajer argues that a discourse, which can exist in documentary form or in speech, is:

> "... an ensemble of ideas, concepts, and categories through which meaning is given to phenomena" (Hajer, 1993, p.45)

Through imparting meaning to phenomena, discourse is the means by which problems are defined (or framed) and subsequently the nature of the evidence that will be sought to provide policy solutions and the way in which that evidence is interpreted in the course of making policy (Hajer, 1993; Fischer, 2003; Hajer, 2003; Considine,

2005). It is because the deliberative process in policy making is not simply the exchange of facts, values, or theory, but rather a process of argumentation that inherently consists of all three, that it is so important to analyse the discourse of these deliberative processes; discourse is the essence of what policy makers do (Hajer and Wagenaar, 2003).

Working towards an understanding of the dynamic nature of discourses is an important part of their analysis, for it is in the coalitions that form (through discourse) between different actors that the policy making process may be better understood. Coalitions typically form around discourses that comprise of a number of other discourses drawn from different fields, for example from the scientific, economic, and political domains (Hajer, 1993). It is because the complexity of these discourses is hidden by the formation of a coalition that it becomes so important to analyse the discursive construction that lies below the surface, for it is this which fundamentally structures the definitions and understandings by which policy making proceeds (Hajer, 1993; Fischer, 2003). In social policy, this becomes more important still, given the wider cultural narratives that inform the construction of what are considered to be social problems (Hajer and Wagenaar, 2003).

The task of analysing discourse that is used in policy making centres upon the way in which discursive constructions (shared ways of conceptualizing and deliberating upon a problem) are brought into existence and utilized (Fischer, 2003). The key issue here is how issues are framed: Why are certain features problematized and not others? How are the actors positioned so as to be victims or causes of the problem? How does the framing direct attention towards certain types of solution? (Fischer, 2003; Considine, 2005). This framing of the issues is crucial for establishing the discursive framework in which policy makers deliberate; indeed, it could be argued that deliberation could not take place without it. What is at stake, however, is an awareness

of the subjectivity and particularity of the framing, as without this awareness one can be under the illusion that policy making is conducted in an entirely objective manner.

Hajer (2003, p.103-107) suggests that the policy making process may be best understood as a process of 'mutual positioning' in which the actors in the policy making process actively negotiate, through shared discourses, policy outcomes that are to their satisfaction. These discourses comprise of a number of factors that also serve as useful conceptual terms for the task of analysis (represented diagrammatically in Figure 3). The strength of Hajer's discourse analytic approach lies in the way that it facilitates an analysis of the policy making process that is both cognizant of the wider political-economic environment *and* which allows for the theory-laden nature of the observations (see Chapter 1, sections 1.1 and 1.4) that inform the deliberations of policy makers. This has significant implications for policy making in the era of EBP. If it is indeed the case that policy lenses are of substantive import in the interpretation of 'evidence', and that these interpretations (in the form of policy) feed back into the construction of story lines[22] that frame future efforts to construct and deliberate upon the 'evidence' (see Figure 3), then an understanding of how this occurs in particular areas of policy making is vital.

[22] There is a danger here that the proliferation of novel terms introduced by Hajer may serve to obfuscate rather than clarify the analysis. Table 1 defines the key words used in Hajer's analytic schema and provides a summary of how these constituent parts may combine in the policy making process.

Figure 3: The policy making process (based on Hajer, 2003)

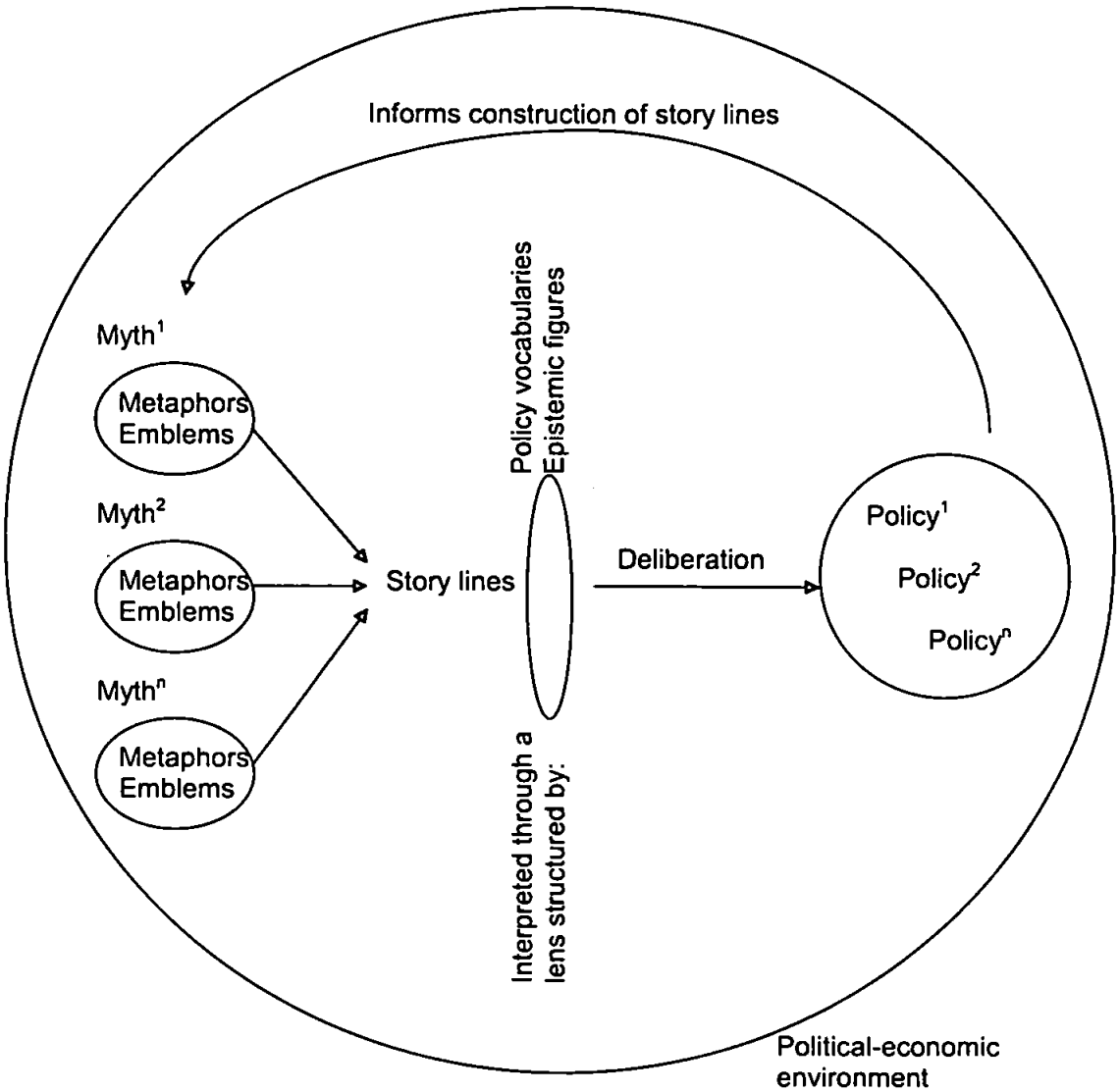


Table 1: Definitions of key discourse analysis of policy concepts (Hajer, 2003)

Metaphors	"... vehicles for the discursive reduction of complexity [that allow] people to communicate over complex policy issues." (p.105)
Emblems	Key figures or representations that stand for the wider problem, e.g. the otter standing for the wider problem of nature conservation (p.105)
Myth	A constitutive or dystopian explanation of why things are the way they are (p.105)
Story line	The contemporary policy agenda that allows different pieces of knowledge to be fitted into "... the larger jigsaw of policy debate." (p.104)
Policy vocabulary	"[A set of] concepts structuring a particular policy, consciously developed by policy makers... [that] determine what was a legitimate policy action and what was not." (p.105-106)
Epistemic figure	"... a regularity of thinking of a particular period, structuring the understanding of reality without actors necessarily being aware of it." (p.106)

A discourse analytic approach to the policy making process takes account of the manner in which *metaphors* and *emblems* are bound together by *myth* and integrated into the contemporary topic of concern through *story lines*. These story lines are interpreted through a policy community lens that is structured by *policy vocabularies* and *epistemic figures*. Policy is formed through the deliberation of (a range of) 'evidence' within this frame, which structures both how the deliberation will take place and the admissible forms of evidence (after Hajer, 2003).

It is by utilizing Hajer's schema for the analysis of the policy making process that the history of drug policy making in the UK since the mid-nineteenth century will now be presented.

2.0 Drug policy making in the UK

The analysis in this chapter focuses upon four key periods in drug policy making in the UK; the trade and regulation of opium by the British in the nineteenth century, the formation of the 'British System' in the early twentieth century, the post-war decades leading up to the Brain Reports (1961 and 1965), and the 'heroin epidemics' and growth of the harm reduction approach in the 1980s. These have been selected for the manner in which they demonstrate the complexity of the policy making process and for the insights that may be gained regarding the uses of argumentation and the ebb and flow of different discourses surrounding substance misuse. The analysis of these periods of drug policy making is made in an effort to give greater insight into the contemporary framing of substance misuse. The deepest analysis is made of the

genesis of policy regulating opium in the nineteenth century; this is because of the degree to which contemporary framings are rooted in conceptualizations of the issue formed in this era.

2.1 The trade and regulation of opium by the British in the Nineteenth Century

From a twentieth century Western perspective, the manner in which opium was consumed in the UK during the first half of the nineteenth century appears extraordinary. That a currently illicit substance was available at a reasonable price from almost any general shopkeeper and was utilized in a range of home remedies for everything from soothing restless infants to relieving the maladies of poverty (whether from toothache, or the diarrhoea or coughs that resulted from insanitary and overcrowded urban conditions) (Berridge, 1999)[23] seems almost fantastical. Furthermore, to the modern eye, the fact that the greatest concerns voiced over opium were those regarding its price and quality (opium being viewed no differently to other traded commodities, among them tea, tobacco, and sugar) sits uneasily beside contemporary understandings of opium as a substance capable of fuelling great social malaise. The question is thus raised of how the currently dominant understanding of opium, and by extension that of other substances that have been classified as illicit, came to be. This section will therefore trace the confluence and synthesis of different forces that established the regulatory framework that has in essence persisted until the present day. In doing so, the manner in which this framework was borne of a social and political system (in conjunction with a developing scientific system) rather than from a more detached assessment of the dangers posed by the substances concerned, will be traced.

The unregulated distribution and sale of opium and its preparations (such as laudanum - opium dissolved in alcohol and distilled water) in the UK during the first half

[23] Virginia Berridge is widely acknowledged as the leading authority on the early development of drug policy in the UK.

of the nineteenth century was such a part of everyday life in both the working and middle classes that it warranted little attention. Whilst it was acknowledged that the use of the substance could be deleterious, the idea that it should be subject to regulation simply did not exist. The recognition that the deaths of some infants were attributable to large doses of opium containing preparations[24], and that these preparations were sometimes used in order to commit suicide (Berridge, 1999), was not manifest in a form that could be considered to be emblematic; they were simply regrettable instances that were part of day-to-day life. As such, there was no story line to weave together in order to call for policy action, the general perception simply being that the use of opium was not a problem.

By the middle of the nineteenth century, the growing influence of the medical and pharmaceutical professions meant that infant deaths attributable to opium overdosing became a key emblem within a story line that posited a myth about the dangers of opium preparations. The unregulated sale of substances thereby came to be discussed as a public health issue, where before it had simply been a non-issue within the domain of the individual and their family. As is made clear by Berridge (1999), humanitarian concern for consumer safety and the professional self-interest of the medical and pharmaceutical professions happily coincided in their pursuit of domains of expertise that would be enabled through regulation[25]. This led to the passing of the Pharmacy Act (1868), which restricted the sale of opiate preparations (in bottles labelled 'poison') to pharmaceutical chemists, who were also required to record all sales, the purchaser, and the purpose for which the substance was required (Mott and Bean, 1998). The medical profession was not bound by this Act, and doctors maintained the right to directly supply drugs (including opiates) to their patients. This was a key moment in the history of the regulation of substances in the UK; even if, as

[24] Such preparations were routinely used to quieten infants; the differing strengths of opium preparations that were nominally the same is one reason why accidental overdosing could occur.
[25] In delineating areas of professional practice and expertise, it was necessary for doctors to assert control over the right to prescribe (self-medication undermines the right substantively) and for pharmacists to assert control over the right to prepare and sell opium containing preparations.

was probable, the Act was widely flouted in practice, opium had been established as a matter of professional concern rather than one which was in the domain of the personal or familial home remedy (Berridge, 1999).

In analysing drug policy, it is imperative to take account of Britain's position as a colonial power in the nineteenth century and the significant role which it played in the opium trade. Opium was exported from British-administered India to China as a means of tying China into a trade network that ensured the continued supply of primary goods from the colonies to the UK (Chung, 1978; Bello, 2005). This 'triangular trade' allowed profits from the sale of UK manufactured goods in West Africa to be used to purchase slaves to work on the plantations of the West Indies. Sugar and cotton produced on these plantations were shipped back to the UK to satisfy consumer demand in a burgeoning economy that had developed a taste for these commodities. China did not fit into this system; it already produced sufficient quantities of commodities such as sugar and tea domestically, and an extensive textile industry meant that there was no need for cotton clothing produced in the UK. However, the Chinese habit of opium smoking remained unexploited. One interpretation of this period was that the colonial power oversaw the unofficial importation of opium from British-India as a means of bringing China into the system of international trade that both provided a market for goods produced in British-administered areas and a cheap source of the stimulants (sugar and tea) that fuelled the labouring of workers in the UK (Trocki, 1999; Bello, 2005).

Care needs to be exercised in the analysis of Britain's role in the opium trade. There is a risk of promulgating stereotypes if China is simply portrayed as the victim of a colonial power that flexed its considerable military, political, and economic muscles in order to attain its economic goals. In this analysis, Britain's policy was wholly cynical; opium constituted just one element of a wider trade strategy that solidified new political-administrative and economic relations. It was argued that these changes substantially

benefited the colonial power whilst the deleterious effects of opium use upon the Chinese populace weakened any ability to resist the new order (Trocki, 1999). Indeed, this has been the dominant historical approach to the trade throughout the latter half of the twentieth century (for example, see Fairbank, 1978), and it is only in recent years that a body of historical work that questions such straightforward analyses has been produced (Newman, 1995; Baumler, 2001b; Dikotter et al., 2004; Bello, 2005). Whilst identifying a common point of departure for these recent analyses risks over-simplification, it may be posited that what they share is an approach that seeks to understand the phenomenon of opium smoking in China within its social context. Opium has been used in China, albeit not always on a wide scale, for 'medicinal' purposes since at least the ninth century (Gray, 1990). It was during the seventeenth century, at least a century before overseas traders started to import opium, that the smoking of domestically-produced opium for its restorative and euphoric effects began to spread in popularity (Blunden and Elvin, 1983).

Opiate consumption in China arguably had considerable similarities to that of the UK. In an era before modern pharmaceuticals, its analgesic, anti-diarrhoeal and cough suppressant properties played an important medicinal role; in addition, the sense of well-being which its use could impart ameliorated the unpleasant sensations of hunger and cold that were frequently the lot of the working class (Dikotter et al., 2004). Arguably, the use of the substance was not the result of an unsuspecting populace being enslaved by the psycho-pharmacological effects of a drug, but rather took place in the context of a dynamic social system that both promoted and regulated its use (Newman, 1995; Dikotter et al., 2004). Newman (1995) details the widespread nature of opium use in China throughout all sections of the society in the nineteenth century; its use by businessmen as a social lubricant in negotiating deals, as a means for those in public office to relieve their professional anxieties, and for members of all strata of society in socializing and relaxing. The consumption of opium was socially

regulated, both by the patterns of consumption inherent to the social occasion and the risk of public censure for overindulgence or solitary use (Dikotter et al., 2004).

To highlight the social nature of opium consumption in China is not to suggest that the use of opium was wholly beneficial. However, the attribution of widespread social malaise to its use is highly problematic given the difficulty of separating cause and effect. Was malnutrition in poor farming communities a consequence of the sums of money spent on opium or of the marginal conditions in which people lived? The attribution of causality is made more problematic still by the paradoxical nature of opium's effects on the body; whilst its use can induce sleep, its analgesic and euphoric effects can inure the body to physical stress and allow hard labour to continue. The habitual smoking of opium is therefore not necessarily incompatible with productivity[26]. Newman (1995) argues that to focus attention solely upon the deleterious impacts of opium for some users is to miss the point. This is not to deny the inherent properties of habitual opium use (tolerance and physical dependence), but to question what the real impacts upon Chinese society were. Most pertinently, how was it possible that in a society where opium consumption was widespread, where the pernicious effects of the drug supposedly spread through the population in a manner akin to a biological epidemic (Baumler, 2001a; Dikotter et al., 2004), there were many Chinese who did not smoke opium in a habitual fashion? (Newman, 1995)

The perception in Britain of the effects of habitual opium smoking in China substantively fed into the development of drug policy. To understand how the concept of addiction developed, it is necessary to understand the roles of not only the medical profession, but also the Society for the Suppression of the Opium Trade[27] (SSOT). European missionaries were the primary source of knowledge about China in the

[26] For example, see the ethnography of opium use in Rajasthan, India by Ganguly (2004); the integration of habitual opium use into the socio-cultural life of rural communities was found to act as an important means of sustaining productivity rather than jeopardizing it.
[27] The SSOT was formed in 1874 by a group of Quaker reformers; Methodists, Presbyterians, and Unitarians also joined the campaign.

nineteenth century (Baumler, 2001b)[28], and it was the reports of these missionaries that the smoking of opium was terminally destructive both to the physical and moral self *and* to wider society that substantively informed the SSOT's work[29]. The SSOT drew strongly on the manner in which missionaries had 'framed' the smoking of opium. This informed the promotion of a discourse about opium use that was strongly located in Quaker beliefs about righteous behaviour and its effect upon the soul, where opium was attributed with the power to cause not just physical addiction, but also to deleteriously affect the moral fibre of the user (Harding, 1988). This melding of medical and religious knowledge (termed by Harding as the 'moral pathological' model of opiate use) came to be integrated into the medical discourse of opiate use as a disease in the form of a scientifically-established fact, even though its roots were substantively located in Quaker understandings of righteous living (Harding, 1988; Berridge, 1999).

The moral pathological model of opiate use is highly significant; a concept of addiction was woven together from metaphors rooted in a particular religious conception of the world and myths about the deleterious impact of opium use on the economic development of China, which then assumed a scientific authenticity through the medical profession's utilization of it. It was a short step from this concept to the framing of the use of opium as a social problem; users, already impoverished by financing their addiction, were furthermore incapacitated from exercising moral judgement (that is, to see how their behaviour was also impoverishing their dependents) due to the effects of opium use (Dikotter et al., 2004).

The development of the concept of addiction was not isolated from the wider international context. The SSOT was successful in promoting the moral pathological model of opiate use at least in part because of the middle-class fears that could be

[28] For example, see the extract given by Baumler (2001b, p.35-42) from Justus Doolittle (1865) *Social Life of the Chinese*. New York: Harper.
[29] Although missionaries were not uniformly critical of the practice; in particular, Roman Catholic missionaries in China tended towards the more moderate view that smoking opium was a social practice that should be tolerated. However, as missionary zeal intensified towards the end of the century, a more critical perspective began to be adopted (Dikotter et al, 2004).

played upon regarding the 'exotic' use of opium. Racial prejudice in the portrayal of opium dens in Chinatowns in the UK as places where young white girls were tempted into sexual liaisons with 'sinister Orientals'[30] were not uncommon (Dikotter et al., 2004). Such portrayals were interwoven with more humanitarian, perhaps even paternalistic, concerns for the role that Britain's opium trade was playing in fostering (or harming) the development of the people of China (Berridge, 1999). This was further wrapped up in an evolving discourse of modernization (in Hajer's term, an epistemic figure of the period), whereby the negative association between using opium and being a useful (economically productive) member of society was stressed (Smart, 1984).

The various facets of the moral pathological concept of addiction thus proved popular amongst sections of the elite in the UK. The Chinese political elite were also able to utilize this discourse surrounding the nature of addiction to opium as a means of pursuing a number of their own goals. It is important to be aware that the power of this elite depended in no small part upon the goodwill of provincial members in exercising local control and collecting taxes; in turn, all were economically-bound into both the colonial and traditional trade in opium (*inter alia*, within China itself, Kashmir, Badakshan, and Russian Central Asia (Bello, 2005)). To shore up power, it was necessary to ensure that opium could continue to be traded extensively, whilst simultaneously marking out China's role in a rapidly developing global mercantile economy (which required a willing and fit workforce) *and* portraying the colonial power as a destructive force that was foisting an addictive substance onto the Chinese. For their part, the Chinese discourse was also overlain with notions of racial purity and the idea that white colonialists were sowing the seeds of racial extinction in the Chinese with the use of opium (Dikotter et al., 2004).

[30] The prevalence of (white) British fears regarding miscegenation and how substance use by non-whites and young women in the UK has been portrayed as fostering such an 'undesirable outcome' is analysed in detail (in the first quarter of the twentieth century) by Kohn (1992). It does not seem unreasonable to argue that similar fears about miscegenation were prevalent around forty years earlier, especially when the extent of xenophobia towards Chinese immigrants to the US in the second half of the nineteenth century is considered (Musto, 1999).

The strategy was successful for both the Chinese and the British governments, insofar as an article added in 1885 to the Chefoo Convention of 1876 limited taxation of opium in China to a single import duty. This article gave the Chinese government a greater share in the profits of the opium trade whilst integrating the country further into the global mercantile system *and* keeping the domestic consumers of opium supplied (Berridge, 1999). It did, however, significantly weaken the SSOT's argument that Britain was forcing opium onto the Chinese; it could hardly be argued that this was occurring when China had agreed to a Convention where the country benefited financially from the continuation of the opium trade. The SSOT subsequently splintered into a number of groups that held differences of opinion on how the opium question should now be addressed[31]. The anti-opium movement subsequently faltered even further following an extensive investigation of the 'opium question' in India (published as the Report of the Royal Commission in 1895), which found opium use to be largely unproblematic and hence not in need of further regulation, let alone the discontinuation of Britain's involvement with the opium trade (Berridge, 1999).

Whilst the question regarding Britain's involvement in the opium trade in the second half of the nineteenth century was primarily framed by missionaries and the SSOT, domestically it was the medical profession (utilizing the moral pathological model) that held sway over the discourse surrounding opiate use. Notably, given the concern voiced at various points in the twentieth century, this meant that opiate use by the working class was only considered problematic inasmuch as the accidental poisoning of infants and the 'over-use' of opiate solutions to quieten infants were regarded as public health issues (Berridge, 1999). The medical profession's primary concern was with respect to its middle class clientele[32], in particular those who had

[31] Among these groups were the Christian Union for the Severance of the Connection of the British Empire with the Opium Traffic, the Women's Anti-Opium Urgency League, and the Anti-Opium Urgency Committee.

[32] Aside from the perception by working class opiate users that the habit and its effects were entirely unremarkable and hence not an issue about which a doctor would be consulted, the majority of the working class lacked the financial means to pay for a medical consultation. The medical profession's

become dependent upon morphine (administered subcutaneously) after being medically prescribed the substance for 'nervous conditions'. This opiate use was not widespread, yet the fact that it constituted the majority (if not all) of the medical profession's experience in managing opiate dependence meant that it dominated the development of the concept of addiction.

Arguably, addiction was not so much 'discovered' in the late nineteenth century as 'created' (Parssinen and Kerner, 1980). The medical profession was buoyant following the growing influence of the (medical) germ theory of disease, and was keen to consolidate its increasing professional power by clearly delineating its areas of expertise. Whilst the profession had been aware of the effects of opium (dependence, tolerance, and withdrawal) since the eighteenth century, the definition of addiction required a number of drivers outside of the straightforward observation of patients. Arguably, the creation of the concept of addiction was fuelled first by a desire in the profession to demonstrate the ability to self-regulate; this was seen as becoming increasingly problematic when physicians were unquestioningly prescribing morphia to their patients. Second, phenomena were increasingly being redefined in scientific, medical terms[33], thereby granting the medical profession an important role as moral guardians (Parssinen and Kerner, 1980). Debate within the medical profession over the effectiveness of treatments for withdrawing from opiates[34] played a crucial role in affirming the profession's belief that it was developing an expert body of knowledge that justified, even required, its leading role in the management of addiction (Berridge, 1999).

---

knowledge was thus almost entirely based upon its experience with the middle class users who consulted them.

[33] For example, madness was redefined as mental illness, and drunkenness as alcoholism (Parssinen and Kerner, 1980).

[34] Treatments that were advocated for the treatment of opiate withdrawal (none of which were necessarily mutually exclusive) included the pharmacological (potassium and sodium bromide, bromide and caffeine, cannabis, coca, and even the newly-synthesised opiate, heroin), psychological (hypnotism), and moral ('re-education' and 'restoration of the will', frequently through religion). In addition, 'wholesome activities' such as exercise and bathing were advocated (Berridge, 1999).

A double-standard was in operation that depended upon the status of the person taking opium and the route by which it was administered. A dystopian myth regarding the 'failure' of China to develop was largely explained by reference to the emblematic Chinese peasant who smoked opium and subsequently suffered the destruction of his moral fibre[35]. However, the injection of opiates by the (white) middle class clientele of the medical profession in Britain was notable by its absence in terms of emblems, for the practice was confined to private spaces and held to be confidential within the doctor-patient relationship. Whilst the moral pathological model was still held to apply in these instances, there was no concern that the practice would spread deleteriously throughout society in view of the 'respectable' nature of the users and the story line which held that opiate use was (in certain cases) acceptable under medical supervision.

Table 2 provides a summary, using Hajer's framework, of the development of British drug policy in the nineteenth century. The extent to which a conceptual framework regarding substance misuse was developed and solidified in this era is of substantial importance for understanding contemporary drug policy making, especially when the evidence-base for the framework is of such dubious provenance. If the foundations of contemporary conceptual frameworks regarding substance misuse were laid during this era, what informed their construction? Table 2 shows how drug policy began to be developed based not so much upon an assessment of the intrinsic risks associated with particular substances as upon a multitude of factors external to the substances themselves. In this way, Britain's self-perception as a colonial power for global economic development intermingled with xenophobia and a work ethic (emphasizing sobriety) that was applied to both colonial peoples and the working class of Britain. The pursuit of areas of expertise by the medical and pharmaceutical

[35] Whilst women in China also smoked opium they did not do so to the extent of men, nor was their opium use publicly visible in the same manner as was men's use in opium dens.

Table 2: Discourse analysis of British drug policy in the nineteenth century

Political-economic environment	Colonial Britain largely pre-eminent as a world power. Colonies viewed as a source of commodities, but also as peoples who required moral and economic development. Colonies viewed as requiring integration into a global mercantile system.
Metaphors	Diminution of a person's moral fibre through the use of substances (e.g. alcohol, opium), in the same manner that a person could be corrupted by evil.
Emblems	Emaciated Chinese peasantry (those who smoked opium). Down-at-heel opium dens full of soporific Chinese men. Chinese men, whose dubious morality was intensified by their smoking of opium (in particular, Chinese opium smokers in Britain who 'took sexual advantage of' white women).
Myths	'Failure' of China to rapidly develop economically attributed to a morally weak people whose lack of adherence to routine work was worsened by an addiction to opium. Opium exerted too strong an effect upon the individual to allow them to self-regulate use – if deleterious consequences were to be avoided, professional supervision was required.
Story lines	Economic development (in the interests of all) is founded upon the work of a disciplined citizenry. It is useful and valid for humans to use substances that stimulate (coffee, tea, sugar) but not those that sedate (alcohol, opium), except under medical supervision.
Policy vocabularies	Addiction (which required professional intervention). Public health (the right of the government to enforce regulation in areas previously in the private domain). Germ theory of disease (contagion). Moral pathological model of opiate use.
Epistemic figures	The global 'enlightenment' project – a person's place in society is dependent upon their being economically productive rather than their social position at birth.
Dominant actors	Medical profession. Pharmaceutical profession. Society for the Suppression of the Opium Trade. Missionaries.

professions, and a growing body of public health bureaucrats, fuelled the development of a moral-pathological conception of addiction that legitimated these professions' interventions. This knowledge was based primarily upon experience gained treating middle class clientele and involved the enshrinement in scientific language of a moral approach to substance misuse that substantively drew upon particular religious understandings of the effects of certain substances upon the soul.

2.2 The early twentieth century (1900-1926) and the formation of the 'British System'

The first decade of the twentieth century was marked by a lack of concern in the UK regarding substance use. The Poisons and Pharmacy Act (1908) changed little substantive from the regulatory framework set up by the Pharmacy Act (1868) and few concerns were voiced publicly with regard to the use of opiates or any other substances (Berridge, 1999). However, significant changes were afoot with the dramatically increasing influence of the United States of America (US) upon international drug policy. This influence played a key role in the formation of what has become known as the 'twin-track' UK approach to drug policy, within which both the medical definition of addiction as a disease and the (US-fostered) penal conceptualization of substance use were accommodated. The development of this approach can be traced through the international conventions that were held, the legislation that was passed in both the UK and US, and the recommendations of the Rolleston Committee (Ministry of Health, 1926); the degree to which domestic US concerns that portrayed a causative link between substance use and criminal activities (most pertinently by ethnic minorities) led the formation of international and UK drug policy is argued here to be substantial, and moreover of great importance for developing an understanding of contemporary drug policy making.

The seizure by the US of the Philippine Islands from Spain in 1898 resulted in the country having a new economic and political presence in the Far East that was perceived to entail responsibility towards the (economic and moral) development of the

islands' people (Musto, 1999). Views on the deleterious effects of opium use on the islands were substantially informed by the reports of US missionaries in the region who had adopted a similar moral pathological model of substance use to British missionaries in China. Keen to establish a sound basis for policies in its new colonies, in 1903 the US government commissioned an extensive review of the manner in which the 'opium problem' was addressed in other Far Eastern states[36]. Bearing distinct similarities to the manner in which British missionaries to China conceptualized opium use, the committee recommended a gradualist approach that moved towards increasing regulation, and eventual prohibition, of the substance[37] (Musto, 1999, p.27-28). The metaphor that a person's moral fibre would irresistibly be denigrated by use of the substance, as promoted in the discourse of the SSOT in Britain, was clearly evident in the report's analysis.

The US was experiencing increasing vigour on the world stage at this early point of the twentieth century and sought to exercise this power in its own economic and political interests. Domestic pressure from missionaries for the US to use its influence to persuade the European powers to discontinue their involvement in the opium trade was growing (Mott and Bean, 1998) and the resulting convention of the Shanghai Commission (1909) established the foundations of an international regulatory framework. Embarrassed to find that the extent of the US's domestic opium regulation[38] was limited to the imposition of a negligible tariff on imports, legislation was promptly secured that prohibited the substance (Musto, 1999). Doing so was largely facilitated by the fear that existed amongst whites of the purported connection between substance use by ethnic minorities and undesirable or criminal behaviour that placed

[36] The committee charged with conducting the review travelled to eight states (Japan, Formosa, Shanghai, Hong Kong, Saigon, Singapore, Burma, and Java) in the course of its investigations. Its members reflected the strategic, medical, and moral aspects of the issue; they were Major Edward C. Carter (US Army Commissioner of Health), Dr Jose Albert (a widely-respected physician), and Bishop Brent (a prominent missionary) (Musto, 1999).

[37] Although there were elements of the recommendations that were not so gradualist; for example, opium dens were to be closed and poppy cultivation prohibited.

[38] The smoking of opium in the US had grown slowly but steadily since the mid-nineteenth century, and was largely attributable to the economic migration of Chinese men who came to work on the construction of railroads (Musto, 1999).

the (white) general public at risk. The myth of the devious Chinese man (linked to opium use) and the violent and/or sexually predatory black or Mexican man (linked to cocaine and cannabis use, respectively) acted as strong emblems that fostered the rapid passage of the legislation (Musto, 1999, p. 31-33 and 43-44). The strong lead that the US took at the conference in preparing reports and pushing for substantially tighter international regulation of drugs was bolstered by the largely supportive position taken by Britain and China (Bruun et al., 1975), leading to resolutions concerning the general suppression of opium smoking and a significant tightening of controls on morphine[39]. Despite the international nature of the conference, it was left to the individual member states of the League of Nations to interpret how best to implement drug control (Bruun et al., 1975).

The position adopted by the US at the Shanghai and Hague conferences arguably reflected a desire to assert political and economic power; it was understood that if the Far East was to be cultivated as a market for US goods, two factors needed to be addressed. First, it was recognized that Chinese immigrants to the US had suffered widespread mistreatment and that this needed to be 'set right' in the eyes of the Chinese by policy that overtly demonstrated US concern for the well-being of the Chinese. Second, the colonial activities of Britain needed to be cast in an unfavourable light so that the former colonies might be better disposed towards the US as a trading partner (Bruun et al., 1975; Musto, 1999). Stringently regulating, and moving towards prohibiting, a substance (opium) that was understood to be inherently damaging to not only individuals but also the very fabric of society, was one way in which these complex geopolitical goals could be surreptitiously attained (Musto, 1999, p.29-32).

[39] The resolutions were not implemented immediately; the Hague Conference (1911) clarified the nature of control of proscribed substances (opium, morphine, heroin, cocaine and derivatives were limited to medical uses; preparations containing <0.2% morphine or <0.1% heroin or cocaine were exempt), but it was not until 1914 that the convention was put into effect amongst those states that had signed (Serbia and Turkey refused).

Meanwhile in Britain, involvement in the First World War was fostering an environment that was highly suspicious of foreigners and in which the populace was apt to accept with little questioning any legislation that was presented as being in the interests of the national war effort (Berridge, 1980; 1984). Newspapers played an important role in framing stories relating to substance use. For example, young women who were 'seduced' into associating with foreign performers in the West End of London (and subsequently suffering the, sometimes fatal, ravages of cocaine use) were significant emblems of the 'problem' of substance use, as indeed were the smuggling of drugs on British warships and the use of cocaine by Allied soldiers (albeit to a lesser extent). In this environment, the Defence of the Realm Act (regulation 40B), which prohibited the possession, sale, or administration of cocaine by all except medical doctors, pharmacists, and vets, was introduced unproblematically in 1916 (Bean, 1974). This regulation was uncontroversially extended in the Dangerous Drugs Act (1920) to include opium, morphine, and heroin (Mott and Bean, 1998). The effect of these legal redefinitions, whilst preserving the medical profession's control over the immediate use of such substances, was to begin to shift the emphasis of control away from the medical profession and towards the Home Office (Berridge, 1984).

British drug policy had thus shifted substantially within the course of no more than a decade from a position that regulated certain substances through the medical and pharmaceutical professions, to one where they were legally proscribed. The role of the US in the development of this drug policy cannot be argued to simply be through the imposition of power, for the manner in which US policy was itself rooted in analyses of the 'opium problem' that were promulgated by British missionaries and temperance societies has been made clear. Nevertheless, US influence on both international and British drug policy dwindled in the 1920s; the Advisory Committee of the League of Nations tired of US belligerence towards a gradualist approach to prohibition (Bruun et al., 1975), and Britain became wary of the overtly penal nature of US domestic drug

policy[40] (Bean, 1974). The stage was set in Britain for the development of the 'twin-track' approach to drug policy that was to become known as the 'British System'.

Sir Malcolm Delevingne, a prominent civil servant within the Home Office, played a key role in the development of the 'twin track' drug policy that was to characterize the British approach (South, 1998). Delevingne was acknowledged to be a charismatic and persuasive speaker, as well as a politically-astute operator who recognized that the penal emphasis of the approach being advocated by the Home Office was starting to isolate the medical profession. Subsequently, Delevingne and colleagues began to practice with a 'lighter touch'. When doubts began to be expressed as to whether the regulatory framework regarding morphine and heroin remained fit for purpose[41], the pragmatic response taken was to convene a committee comprised almost wholly of members of the medical profession[42]. This was at the very least a tacit admission that the medical profession had some role to play in drug policy making (Berridge, 1984).

The Rolleston committee was charged with reviewing the available evidence in order to provide recommendations on the following: first, whether the medical supply of morphine and heroin to persons addicted to these substances was warranted upon medical grounds; second, with regard to how the medical profession might best limit the use of such substances so that abuse was minimized; and third, to suggest an administrative model that would enable these proposals to be adhered to (Ministry of Health, 1926). The medical emphasis in the committee was significant in view of the penal direction which British drug policy had taken in the previous decade, but perhaps

[40] For example, the Supreme Court in the US ruled that it was not possible for doctors to treat addicts in a manner that was compatible with the law; the user of a proscribed substance was required to cease use, or face criminal proceedings. In 1924, this ruling was bolstered by the actions of the FBN in prohibiting the manufacture of heroin for any purpose.

[41] Concerns were expressed regarding the individual freedom that doctors had to prescribe morphine and heroin; there was acknowledged to be a disproportionately high number of doctors who prescribed opiates for their own habitual consumption, in addition to the high doses that some patients were being routinely maintained upon and the issue of addicts obtaining duplicate opiate prescriptions by visiting different doctors (Ministry of Health (1926), paras. 6-10).

[42] The 'Rolleston Committee', named after its Chairman (Sir Humphry Rolleston, President of the Royal College of Physicians).

more important still were the close professional links between several of the committee members and the Whitehall bureaucracy which ensured a partnership between the two professional groups (Berridge, 1980). It is the nature of this professional network, how it impacted upon the policy recommendations of the final report and the way in which the wider policy environment at the beginning of the twentieth century shaped this policy, which is the focus of the analysis in the following section.

The deliberations of the Rolleston Committee were therefore focused upon a medical view of substance use, but significantly this was still a conceptualization that was rooted in the myths that sustained the moral pathological model. The story lines that were developed focused upon structuring discussion about what interventions medical practitioners should make in the treatment of addicted substance users. This first required that a distinction be forged between the administration of morphine and heroin for 'medical treatment' and non-medical purposes (Ministry of Health (1926), paras. 6, 11, and 13-15). Second, and of primary importance, a policy vocabulary was developed in which the habitual use of narcotics was re-defined as a *disease*[43] rather than the result of moral weakness (Bean, 1974; Harding, 1988). This was significant for the way that it represented the solidification of medical control over the definition and the treatment of the phenomenon; it was only medical doctors who, in exercising their clinical judgement (as distinct from a non-medically qualified person who exercised their own powers of reason), could diagnose addiction and prescribe appropriate treatment (Harding, 1988). The Report provided extensive consideration of the merits or otherwise of different methods of treating opiate withdrawal, and of the disagreements regarding the effectiveness of these treatments within the committee (Ministry of Health (1926), paras. 35-42). In short, the desire of the medical profession to self-regulate and delineate areas of professional practice (which in addiction still largely consisted of treating middle class clientele) fused in such a way that the

[43] Addiction "... must be regarded as a manifestation of a disease and not as a mere form of vicious indulgence... the drug is taken in such cases not for the purpose of obtaining positive pleasure, but in order to relieve a morbid and overpowering craving." (Ministry of Health (1926), para. 27).

committee concluded that the primary role in the supervision and treatment of the substance user should be taken on by those who were medically-qualified.

The emphasis upon the role of the medical profession and its definition of addiction was important for the way in which it brought greater control of the policy making agenda back towards the profession on a number of levels. The medical profession's monopoly over the definition of addiction was clearly important with regard to exercising professional power within policy making, but it was also with regard to individual doctor's relationships with their patients that the recommendations of the committee kept the management of addiction within the medical domain. The administrative recommendations of the committee served to make prescriptions of narcotics 'once only' rather than allowing them to be dispensed indefinitely by a pharmacist (Parssinen and Kerner, 1980), the effect being to bring doctors and their addicted patients (the majority of whom were middle class women addicted to morphine, administered subcutaneously) closer together in the management of the 'condition' (Bean, 1974; Berridge, 1984). The committee also objected to a system of notification of addicts, largely on the grounds that it would destroy the confidentiality of the doctor-patient relationship (Ministry of Health (1926), paras. 80-82). Moreover, the committee's recommendations consolidated the power of the medical profession to regulate its own members (Berridge, 1984). The committee was emphatic in its rejection of the Home Secretary's powers to deal with the misuse of opiates by medical doctors (whether prescribed for personal or other's use) through the legal system (Ministry of Health (1926), paras. 67-75).

The above analysis should not be taken to imply that the Rolleston Committee was unanimous in its decisions. The role that should be played by maintenance treatment and the notification of addicts was contentious, although there was near unanimous agreement on the concept of addiction as a disease (prison doctors only being inclined to disagree on this point) (Berridge, 1980; 1984). It did however

represent a strong claim by the medical profession for its role in drug policy making, although the portrayal of this as a victory for medical science over the reactive policy prescriptions of the Home Office is untenable (Berridge, 1984); indeed, it might even be argued that given the small scale of the problem of narcotic addiction at this time, the Rolleston Report "... posed a solution to a problem when there was no problem there in the first place" (Bean, 1974, p.68). Arguably, what the Rolleston Report best represented was the recognition that drug policy making could not simply be decided by a single professional or State body. The 'British System', where drug policy accommodated both penal and medical approaches (Berridge, 1984), was thus set in motion and was to characterize British drug policy for the remainder of the twentieth century.

Table 3 provides a summary, using Hajer's framework, of the development of British drug policy in the early twentieth century. The moral pathological model of substance misuse that was developed in the nineteenth century substantively framed this policy making process in the context of significant developments on the world political stage. The strident nature of US drug policy (itself informed by a particular interpretation of the moral pathological model that was interwoven with racist and xenophobic attitudes towards blacks, Mexicans and Chinese) substantially directed the increasing penal emphasis in the developing international regulatory framework. In Britain, this was accommodated within a medical model of addiction in the form of emphasizing the role of *prevention* that the regulatory framework should play; increasing concern over the 'uncontrollable' nature of substance users fuelled this development. It was perceived that, if managing substance use was problematic outside of the physician-middle class clientele relationship, then controlling the production and distribution of illicit substances through a legal framework was the logical option. The arguments for doing so were couched within terms that prioritized

Table 3: Discourse analysis of British drug policy in the early twentieth century (1900-1926)

Political-economic environment	US able to exert strong influence upon international policies, and willing to do so in order to assert economic and political power. First World War gave rise to an environment where emergency Acts could be easily introduced.
Metaphors	Diminution of a person's moral fibre through the use of certain substances (e.g. opium, cocaine), but not others (e.g. tobacco, alcohol).
Emblems	(From the US) 'Out of control' ethnic minority drug users – blacks raping white women after taking cocaine, Mexican prisoners becoming violent after using cannabis, Chinese 'deviousness' fuelled by the use of opium. Deaths of high-profile, 'respectable', women through cocaine overdoses. Smuggling of illicit substances into Britain on British warships.
Myths	Substance use outside of medical supervision was a dangerous affair, particularly for certain groups in which it would cause violent and other criminal behaviour (some ethnic minorities and working class male youths) or behaviour 'unbecoming' of young women (associating outside of mainstream society, and in particular the 'risk' of miscegenation). The 'native population' of Britain required protection from the increasing numbers of foreigners in the country, and in particular the illicit substances that they brought with them and which they encouraged others to use.
Story lines	The necessity of developing means to *prevent* the use of certain substances in order to protect the general public. Development of knowledge within the medical profession so that it could most effectively treat the disease of addiction.
Policy vocabularies	Addiction as a disease. Confidentiality of the doctor-patient relationship. Right of medical profession to self-regulate.
Epistemic figures	Populace expected to defer to the expertise of a burgeoning professional class.
Dominant actors	US geopolitical policy (on the international stage), rooted in US missionaries' analysis of the Chinese and Filipino 'opium problem'. Medical profession. Home Office.

the well-being and moral development of peoples around the globe, but the evidence for the effectiveness of a legal framework for attaining these goals was tenuous, if not non-existent.

2.3 The quiet before the storm (from 1927 to the 1970s)

In contrast to Britain's 'twin-track' approach, the penal emphasis of US drug policy continued to intensify, most notably under the influence of Harry J. Anslinger at the Federal Bureau of Narcotics (FBN)[44]. Anslinger's approach was characterized by the clarity and force with which he pursued stringent drug enforcement policies that carried high fines and mandatory sentencing of offenders (Musto, 1999). This approach met with the general approval of a succession of administrations and the general US populace[45], and reflected a different interpretation of the dominant conceptualization of addiction as a disease. Whereas in Britain the concept had been used as a means of consolidating medical power and expertise through bringing the addict under close medical supervision, in the US the issue was considered to be best tackled at it roots simply by keeping substances and (potential) addicts as far apart as possible through a policy of strict prohibition (Musto, 1999). As with the role of the British press during the First World War, the US press played its part in identifying the purported role of proscribed substances in criminal activities[46] in publishing stories that were highly likely to have originated from sources close to the FBN (Duke and Gross, 1993).

During the 1930s the US re-asserted its influence on international drug policy through adopting a more measured approach that was less likely to alienate other member states of the League of Nations. International policy bore the clear imprint of the strong prohibitionist stance of the US; for example, through imposing clear restrictions on the manufacture of substances and the creation of the estimates

[44] Anslinger held the position of Commissioner of Narcotics at the FBN from 1930 to 1962.
[45] Although it should be noted that Anslinger's advocacy of the prohibition of alcohol did not meet with the same degree of public approval as was the case for other substances.
[46] Cannabis had become the substance attributed with deleterious transformative powers in the US; stories associated with its use appeared widely in the press, including the Florida youth who murdered his family with an axe, the Texas hitchhiker who murdered a motorist, and the West Virginian man who raped a nine year old girl (Duke and Gross, 1993).

system[47] (Bruun et al., 1975; Carstairs, 2005). Nevertheless, by the latter stages of the decade the frustration of the US at the reticence shown by other member states towards its draconian stance (in particular with regard to US desire to substantially tighten the regulation of cannabis) led to the US delegation walking out of the 1936 Convention (Bruun et al., 1975). It was during the period 1935 to 1939 that a subcommittee within the League of Nations produced a succession of well-considered reports on cannabis, examining (amongst other issues) the claimed characteristics of habitual users and the use of cannabis within different cultures. Disagreement amongst the member states as to the deleterious (or otherwise) effects of cannabis persisted into the 1940s, with the energetic promotion of the 'proven' link between cannabis use and crime by the US jarring uncomfortably with the Mexican and Indian positions that were tolerant of its use (Bruun et al., 1975).

International policy thus continued to be substantively informed by a framing of the issue of substance use in terms of the moral pathological model that was solidified at the turn of the century. The international policy arena was a site in which the different interpretations of the concept of addiction as a disease were played out, with the extreme penal emphasis of the US clashing with various regulatory positions and accommodations that reflected individual member states' own histories. Much to the distaste of the US, cannabis had remained outside of this international regulatory framework. It was not until 1955 that a working paper[48] was seized upon as representing the World Health Organization's position and used to justify adopting a penal approach to cannabis use, despite the highly partial nature of the review[49] (Bruun et al., 1975). Cannabis use was positioned as substantively the same as heroin use; it was argued that its use represented a grave risk to the individual's physical and mental well-being and to the fabric of society as a whole. In locating the risk in this way,

[47] In view of the worldwide supply of drugs significantly exceeding the 'legitimate' demand, the 1931 Convention required that each country provide (binding) estimates of their requirements.
[48] 'The physical and mental effects of cannabis' (1955 E/CN.7/L91), as cited in Bruun et al (1975).
[49] For example, the League of Nations' subcommittee reports on cannabis produced during the 1930s were entirely omitted (Bruun et al, 1975).

echoes of the US interpretation of addiction as a disease are clear; the focus was upon the substance user themselves, and whilst medical intervention might be an avenue to pursue, it was preferable to simply restrict the supply of drugs so that the addict could not obtain them in the first place (Carstairs, 2005). On these grounds, it was considered to be a formality that cannabis should be controlled in the same manner as heroin at the international level (Bruun et al., 1975).

Domestic British drug policy after the publication of the Rolleston Report remained largely unresponsive to the changes that were taking place at the international level. In part this was simply due to the absence of any pressing issues that were linked by way of myth to the production, distribution, or use of particular substances. It was not the case that no action was taken at all regarding proscribed substance; for example, an increase in far Eastern shipping traffic through British ports and a more concerted approach to enforcement were responsible for a five-fold rise in offences relating to opium during the Second World War. However, substance misuse remained a relatively minor issue at this time, with offences relating to both cannabis and manufactured drugs (morphine, heroin, cocaine) remaining steady at approximately 100 and 50 offences a year, respectively (Spear, 1969). It was only during the 1950s that some concern over substance misuse began to be voiced. This occurred when the number of high-status users (those in the medical and related professions) began to fall and the use of cannabis and opium by ethnic minorities[50] and working class 'deviants'[51] began to rise (Spear, 1969; Bean, 1974; Yates, 2002). Even so, the extent of drug use was so tightly circumscribed within these communities, that Spear (1969) was able to trace, in detail, the friendships and associations of approximately half of the opiate users in London and the individual acting as their supplier.

[50] Particularly amongst Indian and Chinese communities and theatrical and musical performers from overseas who were performing in the West End of London.
[51] Whilst middle class users of opiates had long been viewed as 'deviant' in the sense that their substance use was outside the norms of society, the new working class users' 'deviance' was qualitatively different. This was because of the way that these users rejected medical treatment of their addiction, where middle class users broadly accepted its legitimacy (Bean, 1974).

It was against this background that the terms of reference of the first Brain Committee were set to review the recommendations of the Rolleston Report and to advise on additional treatment or administrative procedures. The committee perceived the problem of substance use to be small, static and self-contained; personal use by members of the medical profession was 'infrequent', and the sale of 'surplus' opiates prescribed to addicts was non-existent (Ministry of Health and Department of Health for Scotland (1961), para. 36). The committee's recommendations changed little from the Rolleston Report; the treatment of addiction was to be on the basis that it was a disease rather than an act of deviancy, specialist institutions were not required to deliver this treatment (and nor did these addicts need to be centrally-registered), and the medical profession was to continue to self-regulate (Ministry of Health and Department of Health for Scotland, 1961). This first report has been criticised for its 'hands-off' approach to the regulation of medical prescribing of opiates (Bean, 1974; Spear, 2005) and for its negligible use of the evidence with regard to the changing nature of drug use (Yates, 2002).

A second report was thus required of the committee (Ministry of Health and Scottish Home and Health Department, 1965) only shortly after the publication of the first, in order to review its recommendations in the light of figures indicating what were now considered to be significant increases in drug use[52] (Yates, 2002). It was necessary for the committee to tread a delicate path in their deliberations, as on the one hand it needed to be acknowledged that a substantial proportion of the heroin and cocaine on the illicit market originated from medical prescriptions, even though these did not represent any illegal activity on the part of the medical practitioners concerned (Ministry of Health and Scottish Home and Health Department (1965), paras. 9-13). On the other hand, the quantities prescribed by (a limited number of) these practitioners were so large that a patient's claim that the prescription was solely for individual use

[52] Between 1959 and 1964, the number of people in the UK known to be addicted to heroin and cocaine rose from 68 to 342, and from 30 to 211, respectively. Of particular importance, in view of the committee, was the five-fold increase in heroin addicts in the 20-34 year age group (Ministry of Health and Scottish Home and Health Department, 1965).

seemed highly unlikely. The committee thus acted in order to preserve medical autonomy in the field of addiction; it did not want to drive addicts away from doctors (and hence out of reach of their expertise), but neither did it wish to promulgate the situation whereby doctors could prescribe almost unlimited quantities of opiates. Very importantly, the committee wanted to preserve the medical profession's right to self-regulate.

The second Brain Report therefore recommended the closer management of drug users by doctors, and the tightening of the profession's own regulatory framework. The provision of treatment centres (which later became known as 'clinics') to wean addicts off of their drug use, the power to compulsorily detain addicts in these centres if required, and the limiting of opiate prescriptions to doctors working in the treatment centres, were thus implemented through the Dangerous Drugs Act (1967) (Ministry of Health and Scottish Home and Health Department, 1965; Bean, 1974; Yates, 2002). A statutory requirement for doctors to notify the Home Office of all new addicted patients being treated also came into effect, reflecting the concern that addicts drawn from a wider socio-economic spectrum than doctors' middle class clientele should be subject to monitoring by the Home Office (Bean, 1974).

The changes in the conceptualization of substance use in the second Brain Report are notable for the way in which addicts are positioned. In the earlier part of the twentieth century, whilst the Rolleston Report had acknowledged that the spread of substance use by social contagion *could* occur, this was only on a very small scale and was dwarfed by the problems of addiction following the medical use of opiates for organic disease (Ministry of Health, 1926). Now that there was evidence that substance use was taking place on a larger scale outside of medical supervision, addiction was conceived of in the form of a myth about its dystopian effects in deleteriously affecting

the social fabric[53]. The young, who were tempted by fashion into experimenting with amphetamines, cannabis, heroin, and cocaine in West End clubs, were the emblematic victims of substance use. They informed a story line which strongly advocated that action be taken both in the interests of these victims and in the interests of society as a whole (Ministry of Health and Scottish Home and Health Department (1965), para. 40).

There exists a danger that the recommendations of the second Brain committee are viewed simply as a victory for compassionate and scientific medical care over the legalistic conception of the addict as deviant. However, this period of policy making is arguably a more complex affair, reflecting wider social and technological changes as well as the medical profession's desire to retain status and expertise in the policy making process. These changes first occurred *inside* the medical profession; the growing body of formal knowledge of psychiatry laid claim to a recognition that psychiatrists had a distinct expertise in treating addiction (Smart, 1984), and it followed logically that such distinct treatment should be delivered outside of the 'mainstream' hospitals. Second, novel medical treatments for addiction such as methadone provided a rationale for the close supervision of addicts within the clinics rather than the community, and the development of databases enabled monitoring systems of addicts to be established in the form of the National Addicts Index (Smart, 1984).

Table 4 provides a summary, using Hajer's framework, of the development of British drug policy in the mid-twentieth century. Whilst this era is not marked out by major developments, the continuation of policy making that utilized the framework developed during the nineteenth century persisted. The geo-political significance of the US contributed substantially to the continuation of the prevention focus in the international regulatory framework, despite its stance being markedly more measured than previously. Nevertheless, the role played by the US in promoting a particular

[53] Addiction "... is a disease which (if allowed to spread unchecked) will become a menace to the community." (Ministry of Health and Scottish Home and Health Department (1965), para. 18).

Table 4: Discourse analysis of British drug policy in the mid-twentieth century (from 1927 to the 1970s)

Political-economic environment	(Internationally) US policy influence wavered, then re-asserted with strong penal emphasis. (Domestically) Complacency with regard to drug policy; viewed largely as an area in which doctors would exercise their professional judgement.
Metaphors	The spread of substance use as a form of epidemic that causes both individual and social decay.
Emblems	(US) Cannabis as a substance with dramatic transformative powers, leading to users committing murder or rape. (Britain) Covert cannabis and opium use by Indian and Chinese communities, with unpredictable consequences. Substance use in the West End of London, often perceived to have been introduced and facilitated by overseas performers.
Myths	Substance use could spread (socially) by contagion in the same manner as an infectious disease. The increase in substance use outside of medical supervision (particularly amongst the non-white and non-middle class, and/or the young) was a dangerous phenomenon (for society).
Story lines	Development of the specialized management of a disease that could have severe social consequences.
Policy vocabularies	Substances required to be regulated internationally if goal of prevention to be attained. Addiction as a disease. Confidentiality of the doctor-patient relationship. Right of the medical profession to self-regulate. Regulation of the medical prescription of opiates.
Epistemic figures	Professional class has the moral responsibility (and authority) to manage society for the well-being of all. Professions should expect to account for and regulate the conduct of their members; total self-regulation is not permissible.
Dominant actors	(Internationally) The US. (Domestically) Medical profession. Home Office.

working paper on cannabis (in a bid to promote a similar penal approach to both cannabis and heroin) was significant. Domestic British policy continued to build upon the dominant conceptualization of substance misuse within the moral pathological model, with the Brain Committee's recommendations aiming to bring substance users back into a closer supervisory relationship with the medical profession in order that they could be rehabilitated.

2.4 Heroin epidemics, harm reduction and AIDS/HIV in the 1980s

Drug policy in the 1980s was marked by its dual nature; on the one hand, a strengthened penal response was evident in the tighter enforcement of measures to reduce drug trafficking, dealing, and use (Stimson, 1987), but concurrently there was also the rapid development of a health-oriented harm reduction approach (Berridge, 1993). This apparently contradictory approach may be best understood in the context of the political-economic environment; the UK's strong political links to the US promoted the development of tighter penalties for offences related to substances, but this was set against a well-developed background of harm reduction in the domestic health policy network.

The danger in analysing drug policy in the era of AIDS/HIV is that the ascendancy of the harm reduction approach is straightforwardly portrayed as the victory of humane pragmatism over the medicalization of the treatment of drug users. This portrayal bears striking similarities to past analyses in which the preference for medical treatment of drug users over penal approaches is claimed to represent a victory for the pragmatism of the 'British system'. Policy making in this era is arguably a more complex affair than such analyses suggest. Nevertheless, drug policy fundamentally represented a continuation of the 'twin-track' approach set in motion by the recommendations of the Rolleston Report. In this approach, medical control proceeded within a wider framework of national and international penal policy

(Berridge, 1993) that were themselves rooted in conceptualizations of the problem of substance use that were formed in the latter stages of the nineteenth century.

The early 1980s were marked by significant increases in the number of addicts notified to the Home Office (from 3425 in 1975 to over 12000 in 1984) and the quantity of heroin seized by customs (from less than 50kg in 1980 to over 350kg in 1984) (Berridge, 1993). Moreover, whilst this increase in heroin use was initially confined to those areas with existing drug subcultures, its use soon began to grow substantively in areas where there was little previous history of opiate drug use (Yates, 2002). Policy making arguably came to be informed by the emblematic nature of the 'heroin junkie', undeserving working class individuals whose individualism had become mis-focused. This individualism manifested as a selfish pursuit of pleasure through substance use rather than the 'healthy' individualism (an epistemic figure of the period) that purported to enable people to take control over their lives and be economically productive. Nevertheless, the story lines of policy making that developed were moderated by the presence of a well-developed network of health-oriented policy makers who had largely converged upon adopting a harm reduction approach (Berridge, 1993). In a pragmatic policy response that balanced fiscal, health and political demands, the Conservative government established the Central Funding Initiative in order to fund both voluntary and statutory treatment services outside of the clinics (Yates, 2002). This meant that the locus of power, at least at the level of service-delivery, had swung away from the medical profession[54].

Berridge (1993) notes how there were both public and 'in-house' aspects to drug policy during this era. Understanding these aspects is crucial to understanding how the harm reduction approach was adopted so quickly in a political environment that was ostensibly hostile to such a radical approach. The public aspect of policy

[54] However, consultant psychiatrists still played an important role within the clinics (which were re-configured as multi-disciplinary 'Community Drug Teams').

making, with which there was broad political consensus, emphasized the penal approach; the dystopian myth of 'new' and powerful drugs wreaking havoc in society facilitated the introduction of tighter measures to reduce trafficking, dealing, and use (Home Office, 1985). Concurrently, the 'in-house' policy network that had been developing since the mid-1970s (consisting of not only medical members, but also voluntary agencies, researchers and Department of Health civil servants), emphasized the health aspects of drug policy (Berridge, 1993). This network reflected the balance of professional power and wider social change, and subsequently fostered the development of story lines and policy vocabularies that reflected these changes. It was acknowledged that the widespread use of methadone (rather than injectable opiates) meant that only highly-motivated users were likely to persist with this treatment, which was the mainstay of the clinics' provision. The policy vocabulary also changed from conceptualizing addiction as a disease to the concept of problematic drug use; this resonated both with the individualistic epistemic figure of the period (substance use was substantively within the control of the individual, rather than a disease that was 'caught') and enabled the use of a story line that allowed a multi-disciplinary approach to the care of substance users. In this way, it was possible to involve non-medical professionals and voluntary agencies, on the basis that these were the most effective way to address health and social issues (Berridge, 1993).

The pragmatic policy response to containing the spread of the AIDS virus was thus manifest in the form of harm reduction, the professional networks for the delivery of which were already substantively in place. Harm reduction not only placed priority upon preventing the spread of infection, it also signalled an attitudinal change towards the drug user. Users were 'rational actors' who required education on the safe(r) use of drugs, not moral opprobrium; professional-client relationships were conceived of as being equal, not hierarchical; and whilst safe drug using practices were to be promoted, the ultimate goal was to help the user, step-by-step, towards abstinence (Stimson, 1990). Whether such an approach constitutes a distinct break with the past or not is

debatable. Berridge (1993) maintains that key elements of harm reduction (the education of the individual, the responsibility of that individual for their own health, and the role played by the voluntary sector) are shared with public health approaches regarding hygiene from the turn of the twentieth century. Whilst this historical analysis is useful, it perhaps misses a very important point about the thrust of drug policy at this time; if the dramatic rise in heroin use in working class communities in the 1980s was as a result of a constellation of factors external to the individual[55], why did drug policy focus almost exclusively upon the individual in terms of education and enforcement?

Table 5 provides a summary, using Hajer's framework, of the development of British drug policy in the 1980s. Arguably, dominant conceptualizations of substance misuse rooted in the nineteenth century continued to frame policy making, but with the important exception that there was an urgent public health issue associated with substance misuse (the spread of AIDS/HIV) that overwhelmed any arguments to utilize predominantly penal interventions, despite the close political ties between Britain and the US. Broader contemporary socio-economic developments, in particular that of an abrupt shift from a welfare state to a neo-liberal economic system that emphasized individuality, also shaped the way that drug policy was framed and made. Addicts were no longer conceptualized as individuals with a disease that would be treated by professionals (as might be expected in a welfare state), rather they were individuals who required educating in how to manage their lives so as to achieve abstinence. The echoes of the moral basis of drug policy making of the nineteenth century are clear in this more modern expression of temperance, even if its evidence-base rests upon some quite uncertain ground.

[55] For example, mass unemployment doesn't 'cause' heroin addiction in any straightforward sense; but it does provide a set of conditions in which the euphoric escape from everyday life provided by the habitual use of heroin can take root. Moreover, such an environment provides fertile ground in which social and economic ties associated with using and dealing heroin can be rapidly set up in place of those no longer provided through 'legitimate' employment (Pearson, 1987b).

Table 5: Discourse analysis of British drug policy during the 1980s

Political-economic environment	UK tightly linked on a political level with the US by way of the close Thatcher-Reagan relationship. Neoliberal structural adjustment resulted in swathes of mass unemployment in many working class areas that had traditionally been reliant upon dominant industries. Widespread fear of an AIDS/HIV epidemic imparted an urgency for policy making – central government needed to be seen to be 'doing something'.
Metaphors	The spread of substance use as a form of epidemic that causes both individual and social decay.
Emblems	'Junkies' – intravenous heroin addicts trapped within a downward spiral of social decay. 'Pushers' – evil (often criminal) men who preyed upon the weaknesses of individuals by way of their illegal trade.
Myths	'New' and powerful drugs (along with novel, more risky, routes of administration) have the power to spread prolifically and with grave effects throughout otherwise 'normal' communities.
Story lines	Co-ordination of a broad range of services (health, local government, education, legal) that would teach the addict harm reduction and facilitate a path towards abstinence.
Policy vocabularies	Substance use had the (strong) potential to become 'problematic', rather than being a 'disease' *per se*. Harm reduction – the risk from AIDS/HIV was greater than the risk from substance use itself. Drug prevention – tighter enforcement of measures to reduce trafficking, dealing, and use.
Epistemic figures	Individualism – people held to be responsible for their own well-being, and if they lacked control they could be *taught* how to exercise control over their own lives.
Dominant actors	Psychiatric specialty of the medical profession, although very concerned to not work unilaterally – therefore, developed partnerships with professions allied to medicine and voluntary groups. Central government (drug control had become a highly sensitive political issue).

3.0 Alternative frameworks for the construction of an evidence-base on substance misuse

The use in this chapter of discourse analysis as a means of unpacking the multitude of factors that have fed into drug policy making in the UK has perhaps resulted in an analysis that lacks the some of the deft touch of the historian's hand. By this it is meant that the scope which a historian has to construct a compelling narrative regarding the global forces that have driven drug production, distribution, and consumption is a little more constrained when using a framework such as Hajer's. Although the use of discourse analysis has allowed the development of policy (and how this has occurred through myths, emblems, storylines and so on) to be tracked in some detail, what has perhaps been lost are some of the broader insights on a global scale that a historical analysis might have provided (for example, Courtwright, 2001). This is not a weakness of Hajer's method, so much as an observation that, just as one has to balance internal, external, and construct validity in research design, so one also has to appreciate that no analytical framework will be comprehensive.

Nevertheless, the analysis of the development of drug policy presented here allows us to ask how drug policy making might be differently (and usefully) framed. To do this is quite different to suggesting that any of these alternative framings are necessarily superior to the currently dominant conceptual framework, although clearly it would be somewhat odd to make these suggestions if the view was held that the current framework was entirely adequate. To ask this question of how different conceptual frameworks might be useful is to lay the foundations for debate about construct validity in research that endeavours to evaluate the 'effectiveness' of interventions aimed at preventing or reducing substance use.

Before suggesting alternatives that could underpin the conduct of a substance misuse SR, it is useful to summarize what the dominant conceptual framework

regarding substance misuse is in the UK. This is the moral pathological model, which shares certain key similarities with the biomedical model of disease in which pathogenic agents are either destroyed, removed or kept from ever coming into contact with the host. With regard to drugs, this leads to a policy framework that focuses upon rehabilitation (where success is considered to have been achieved when abstinence is attained) and prevention (where legal measures are taken to prevent the distribution and purchase of drugs), rather than one where efforts are made to actively manage the risks associated with the behaviour[56] or addressing environmental drivers of mechanisms that play a substantive role in the initiation and perpetuation of substance misuse. In summary, the dominant conceptual framework revolves around the individual and the substances themselves, rather than the wider social environment.

The conceptual framework of the moral pathological model serves to inhibit consideration of other frameworks that arguably should receive critical appraisal. At the broadest level, the rationale for having a distinct category for illicit drugs is questioned; upon what basis is the consumption of opium or cannabis to be considered substantively different to the consumption of caffeine, sugar, or even spices? Ethnographies of non-European communities provide strong grounds for developing middle-range theories that seriously question the power of 'illicit' substances to wreak societal havoc when their consumption is managed through cultural mores (Goodman, 1995; Hugh-Jones, 1995; Sherratt, 1995). At a more focused level (but still maintaining a society-wide view) the structural role played by the environment may significantly influence the risks involved in behaviour associated with drug use (for example, Rhodes et al., 2005), and hence substantively inform the conceptual framework utilized. Finally, at the level of the individual (but maintaining a strong recognition of the wider environment), there exists a whole body of work that analyses substance misuse as a behaviour through which identity is pursued by those who find themselves in

[56] For example, the consumption of alcohol, sugar, and fats, are managed through a combination of cultural mores and health education, as is sexual behaviour.

disadvantaged positions (Burr, 1987; Taylor, 1993; Davies, 1997; Henderson, 1997; Rodner, 2005). The implications of this approach for a conceptual framework regarding substance misuse are significant, for a particular framing of phenomena will substantively influence how evidence is searched for, appraised and synthesized into a body of knowledge in the form of a SR.

The key question is whether the dominant constructs are fit-for-purpose with regard to drug policy. There are strong reasons to believe that before one can even begin to debate the relative merits of studies with regard to their internal and external validity, it is necessary to seriously consider the validity of the constructs upon which much research into substance misuse is founded. In Chapter 2 (section 1.1) the significance of construct validity was discussed. This chapter has provided the background against which a critical re-examination of constructs used in the field of substance misuse (and subsequently in drug policy making) can be made. Operationalizing constructs that are fundamentally shaped by the moral pathological model, without critically reflecting upon the evidence-base that underlies the adoption of such a model, is argued here to be inadequate.

Conclusion

Tables 2 to 5 have presented a discourse analysis of the development of drug policy in the UK. Presenting an analysis in this way allows the concepts utilized in policy making during different periods to be discerned, and for developments in time to be traced more clearly. Despite many developments, UK drug policy has been characterized by an essentially unchanging conceptualization of the issue of substance use since the mid-nineteenth century. This framing is rooted in the temperance movement and missionaries' conception of substance use as an addiction that leads to the moral breakdown of both the individual and the society of which they are a part.

The myths that have prevailed and guided UK drug policy making have arguably been woven from metaphors and emblems that, whilst they might have changed in appearance, remain fundamentally the same. The moral pathological metaphor, promoted by the SSOT and missionaries, has been present since the latter half of the nineteenth century; it has been applied differently according to the social status of those using substances (for example, middle class women's use was viewed as controllable with medical supervision, whilst Chinese or working class use in the UK was seen as requiring penal enforcement), and it has developed to include the concept of an epidemic when substance use by a non-elite group has spread 'out of control'. The use of emblems in the formation of myths has been of significant importance; again, these changed over time from (amongst others) the soporific and untrustworthy Chinese to the violent Mexican, sexually predacious black, or (white) working class deviant. What unites these emblematic figures is that they are always 'other', always outside of the norms of mainstream society. The emblems arguably have a strongly persuasive function in seeking to unite the general populace behind policy formation that seeks to protect the social fabric that these 'others' are positioned as challenging.

The myths themselves have functioned to present dystopian visions of how society will end up if the issue is not addressed in a certain manner. Medical supervision of substance users, and the implementation of prevention policies in the form of penal measures (with the aim of keeping the (potential) addict and the proscribed substances from coming into contact with one another) have been strongly informed by the myth that certain substances are too powerful in their (addictive) effects to allow individuals to self-regulate their use. These myths have drawn upon wider cultural narratives in their portrayal of the destruction of the moral fabric of society by substance use; for example, 'failure' to be an economically productive member of society, the 'risk' of miscegenation, and the 'danger' of becoming involved in counter-cultures have all at various times substantively informed the myths that fed into the drug policy making process.

It is important to remember that the use of Hajer's term 'myth' here differs somewhat from the more common usage that emphasizes the *absence* of evidence[57]. The analysis presented here makes the case for UK drug policy having been informed to a significant degree by cultural narratives (that is, myths in the more commonly used sense), but it has never been the case that these myths have been totally dominant. Throughout the history of UK drug policy making there have always been 'factual' reports (not necessarily resonant with the thrust of the policy field) interwoven into the myths (in Hajer's sense) that have been part of the policy making process. The Royal Commission on Opium's report of 1895 amounted to 2500 pages[58] and highlighted the degree to which the deleterious effects of smoking opium were exaggerated and its medical and non-medical uses so closely intertwined as to make regulation highly problematic (Berridge, 1999). The Rolleston Report drew upon the deliberations that had taken place in the course of twenty-three meetings, at which evidence was heard from a total of thirty-four professionals[59]. Whilst the predominant source of evidence was the testimonies of these experts, publications detailing medical experience in the US were collated in order to evaluate treatments for the withdrawal of opiates. The Brain Committee was notable for the misplaced confidence that its members initially held in their own knowledge of the field, but were obliged to reconsider their first report in the light of further submissions from the Home Office and the Ministry of Health and the Scottish Home and Health Department. In the 1980s, the imperative of addressing the spread of AIDS/HIV required policy decisions to be made quickly; a multi-disciplinary consensus based predominantly upon anecdotal evidence regarding the plausible effectiveness of harm reduction approaches allowed prompt policy formation.

What is of significance is the manner in which there is an inter-relationship between the reports and testimonies utilized in policy making and the metaphors,

[57] These more common usages of 'myth' are: 1) a traditional story, 2) a widely held but false belief, or 3) a misrepresentation of the truth (Oxford English Dictionary).
[58] The report drew upon meetings in London and data from 723 witnesses ('native', civil servants, and missionaries) in India.
[59] Twenty-six of these professionals were medical (including hospital doctors), six were pharmacists, and two were involved with the criminal justice system.

emblems and myths that are integrated into the topic of concern through story lines. Whilst the broad policy making process is now widely considered to be non-linear, the historical discourse analysis of UK drug policy making demonstrates that the route *within* the policy making process itself by which evidence is analysed is not linear either. Wider cultural narratives inform both the initial conceptualizations used in reports and testimonies and the way in which analysis is *framed* in order to inform the policy making process. Whilst the sheer volume of 'evidence' relevant to drug policy making may have become apparent in the latter stages of the twentieth century, it has never been the case that there was a dearth of evidence. Two questions have always been of primary importance, even if they have not always been explicitly addressed. First, in what way should the evidence be framed in order to allow analysis? Second, in what form(s) is evidence acceptable? Arguably, these are questions are central to all social policy making, and whilst the Evidence-Based Medicine (EBM) and EBP movements clearly address the second question, consideration of the first is for the most part absent.

Summary

The concepts that have informed drug policy making in the UK have reflected a particular world view that was shaped by forces unrelated to the substances themselves. In this way, dominant attitudes towards ethnic minorities, women, and the working class have been encapsulated in drug policy that has drawn substantively upon a moral pathological model of addiction that was developed in the nineteenth century. The conduct of policy making in this way has meant that the intrinsic risks associated with the use of particular substances have informed the process only minimally at best; that tobacco is a licit substance and opium an illicit one demonstrate this particularly clearly. The strong influence, and penal emphasis, of US policy on the world stage have drawn UK policy in the direction of a penal model and away from a more sober appraisal of the risks posed by illicit substances and of more informed ways of managing such risks. The EBP movement provides a substantial opportunity

for critically examining the basis of drug policy making in the UK, but to do so first requires a painful re-examination of the constructs that frame the knowledge cumulation and policy making processes.

Chapter 4

Research Methods

This chapter details the rationale for the methods used in this research and the manner in which they fit together in order to attain 'methodological congruence' (Morse and Richards, 2002). The methods utilized must not only complement each other, they must also fit with the ontological and epistemological position adopted by the researcher. To a substantial degree, ontology and epistemology are implicit in the research questions chosen and the methods utilized (Morse and Richards, 2002; Yin, 2003), but it is a worthwhile exercise to explicitly state these positions at the outset in order to clarify the approach taken in the research. In addition to detailing the reasons for utilizing a case study, ethical considerations in the conduct of the research and the approach taken towards data analysis are discussed.

1.0 Background to the research

This research was conducted under the auspices of an Economic and Social Research Council (ESRC) Collaborative Awards in Science and Engineering (CASE) research studentship. CASE studentships differ from a conventional United Kingdom (UK) Research Council studentship in that they involve explicit collaboration, in the form of a joint research proposal (and responsibility for supervising the research) between an academic and non-academic institutional partner[60]. In addition to the scope of the research and a timetable for its completion being formally agreed to, an intellectual property rights agreement between the university, institutional partner, and student is signed. The student benefits substantially from obtaining privileged research access to the institution concerned, and from an enhanced bursary and additional financial support from the institutional partner. The price paid for such benefits is a reduced level of autonomy for the researcher; all decisions in designing and conducting the research are made bearing in mind the formal collaborative nature of the research.

[60] In this research, the institutional partner was also a senior academic who was working in an institutional role.

The studentship under which this research was conducted (PTA-033-2004-00074) was awarded to the current author for 36 months, commencing in January 2005.

The original proposal submitted for this research (filed with the ESRC 01/12/03) envisaged utilizing the drug prevention evidence-base produced by the Health Development Agency (HDA) (Canning et al., 2004) as a case study for pursuing the exploratory question; How is an evidence-base constructed and utilized? However, the process of awarding the studentship and the rapid pace of production of new evidence-bases meant that the supervisory team agreed in the early stages of the research to change the focus of the research. The drug prevention in vulnerable young people evidence-base produced by the National Collaborating Centre for Drug Prevention (Edmonds et al., 2005) was agreed upon as a more up-to-date case study that would still allow the construction of an evidence-base to be researched. Later negotiations that endeavoured to strike a balance between adequately investigating rapidly developing review methods within the timescale agreed to, whilst also keeping to the essence of the original research proposal (RD 4213-4421 16, 17, 20, and 24/03/06) resulted in a further shift of focus, from a case study of a completed evidence-base on drug prevention, to the contemporaneous research of the construction of an evidence-base on substance misuse in vulnerable and disadvantaged youth. This development in the focus of the research was also enabled by the integration of the HDA (the original organizational partner) with the National Institute for Health and Clinical Excellence (NICE) and the fortuitous commissioning of a new evidence-base on substance misuse upon which this doctoral research could focus.

2.0 Research methodology

The following section makes the case for utilizing a case study approach in this research. Subsequently, in order to locate the researcher's approach within a philosophical spectrum, my epistemological and ontological assumptions are outlined

and their relationship to the research methods clarified. Finally, the manner in which research validity was attained is discussed.

2.1 Rationale for conducting a case study

The conduct of any piece of research is shaped by the institutional context that brought it into being. In this instance, the immediate driver of the research was the collaboration between bodies that gave rise to the CASE studentship under which this research was funded. In my role as researcher, I came to the research with agreement having already been made in the research proposal to adopt a case study approach. However, the constructive relationship between the parties concerned meant that there was room for negotiation over the best research approach to adopt. It is important to note that if there had been reason to doubt that, once the realities of conducting the research had become a little clearer, the case study approach was not sufficient for the purposes of research then there would have been considerable scope for renegotiating this aspect of the research proposal.

In the event, the case study approach was deemed to be strongly justifiable as the preferred means of conducting this research. As succinctly stated by Stake (1995), researchers value the case study approach for its potential to provide insights into both the uniqueness and commonality of the phenomena concerned. Case studies are also of significant utility when investigating phenomena over which the researcher does not have control; in short, when it is necessary to research phenomena within their real-life context and where this context is considered to be of substantive importance for gaining an understanding of the phenomena concerned (Yin, 2003).

The selection of the case, or cases, to be investigated in a case study requires careful consideration. In certain ways, this was shaped by the nature of the CASE studentship; a case study of the production of a systematic review (SR) on substance misuse policy was identified in the research proposal as providing a substantive topic

for doctoral level research. With regard to this, it should be noted that the award and running of a CASE studentship requires that the interests of four parties are balanced; the academic and institutional supervisors, the funding body, and the researcher (student) him or herself. As such, there would have been limited scope to, for example, change the focus of the case study entirely to another review (which might have lain outside the area of expertise of the supervisors). However, I argue here that this 'constraint', whilst important to note in the interests of making the research process as transparent as possible, was not really experienced by the researcher as such. The use of the substance misuse review as a case study was justifiable on grounds above and beyond the important fact that privileged access had been obtained for researching this particular case (Yin, 2003); for example, similar review methods are utilized in other areas of public health and social policy, and these could be informed by research upon the process of conducting the substance misuse review.

2.2 Epistemological position

In order to understand the rationale for the use of particular methods in conducting research, it is important that the researcher summarizes their epistemological and ontological position (section 2.3) (Mason, 2002). Stating an epistemological position allows the critical reader to assess whether there is congruence between the theory of how valid knowledge is established (epistemology) and the research methods utilized. At one extreme lies a 'hegemonic' epistemology; a sharp distinction exists between the detached and rational researcher, and the subjects about which (or whom) full knowledge is to be gained. At the other extreme lies an epistemology that 'situates' the production of all knowledge by dint of it being inherently partial and located within power structures (Johnston et al., 2000).

My epistemological position is one that aims to strike a middle ground between the extremes outlined above. It neither attempts to establish a definitive 'reality' through advocating a certain epistemological superiority, nor does it seek to flatten the

epistemological landscape so that all accounts are considered to have parity. Rather, it is an attempt to stake a claim to a relatively modest epistemological position that seeks to maintain scope for considered argument and debate by allowing for different interpretations of the same social reality (Mason, 2002). Accordingly, the importance of understanding the effects of context on the phenomena being researched (as opposed to attempting to remove these contextual influences) forms an important component of establishing valid knowledge. Furthermore, this means that the effect of the researcher upon phenomena is unlikely to be something that can simply be engineered out; in a sense, social phenomena in which the researcher is involved could be considered to be co-constructed by the researcher and research subjects, but this does not mean that the knowledge produced is simply a constructed artefact.

2.3 Ontological position

Ontology endeavours to define the way in which the world must function for knowledge to be possible (Bhaskar, 1978). Following Bhaskar's terminology, is the world like that envisaged by the classical empiricists where it is possible to directly observe discrete phenomena and thereby establish reality? Or is it like that of the transcendental idealists, where the observation of the constant conjunction of events establishes relationships for which mechanisms for theory-building may be imagined? Or is the world like that proposed by transcendental realists, where it is possible to test the postulated mechanisms in order to better know the world?

My ontological position is emphatically shared with the transcendental realists (hereafter simply referred to as 'realism'). I believe that the world functions in such a way as to allow the testing and observation of the mechanisms that link phenomena, and moreover that it is an understanding of these mechanisms that best places science to contribute meaningfully to policy making. This ontological position has two additional distinctive features; first, it is sensitive to the context in which mechanisms operate and hence is not intended to generate law-like statements that apply across differences in

space and time, but neither does it solely deal in unique occurrences of phenomena (Sayer, 2000); and second, it is distinguished from constructivist ontologies in acknowledging that the world (and very importantly, the causal mechanisms that operate within it) exists independently of human perception (Bhaskar, 1978; Sayer, 1992).

How does my ontological position accord with the methods used in this research? The emphasis of the research is upon gaining an understanding of the process of conducting a SR and the utilization of that review for the purpose of making recommendations. The research thus aims to elucidate the mechanisms involved in that process and thereby to further understanding of the phenomena concerned. My position is that this understanding of mechanisms cannot be attained by the use of particular research methods (such as the analysis of interview data or documents) in isolation. Instead, it is necessary to judiciously use a range of research methods with due regard to the context in which the data was produced. This is done in an effort to elucidate the mechanisms concerned and to test out the validity of the interpretations that are made of how those mechanisms operate.

2.4 Research validity

Internal, external, and construct validity in research, as defined within the body of work of Donald Campbell, were summarized in Chapter 2. Clearly, it is also important to consider the balance to be struck in meeting these forms of validity in the research conducted for this case study; research with very high internal validity will have limited external validity; that with high construct validity will have limited internal (in the sense of establishing causation) validity; and so on. What trade-offs were considered regarding validity in the design of this research?

From the beginning, it may be stated that the emphasis of this research was never upon internal validity; research occurring in settings where the researcher has

little or no control over the phenomena of interest provides few (if any) opportunities for the researcher to manipulate the variables concerned in an experimental fashion. Rather, the approach adopted towards validity in this research was to strike a considered balance between external and construct validity. It was considered important to explore and develop constructs in this field, as there is no established research tradition into the process of conducting SRs for social policy making.

However, the research focus was not entirely upon construct validity; its external validity was also considered important for the contribution which the research could potentially make to informing the conduct of SRs in areas of social policy other than substance misuse. It is worth noting here that external validity in qualitative research is distinct from the use of statistical sampling techniques as a basis for making inferences to a larger population; in qualitative research, it is the coherence of the study and its ability to illuminate aspects of phenomena that provide the basis for making inferences (Gomm et al., 2000). For example, Wolcott (1995) contends that ethnographic research is too context-specific to allow 'generalization' in this sense. Whilst Wolcott's position is tenable to a certain degree, it is unhelpful for facilitating the design of research that allows judicious interpretation and generalization of its findings. To this end, this research was designed in such a way as to highlight the ways in which the phenomena of concern in the case study were likely to be similar or different to those in other areas of social policy (Schofield, 2000), although this approach is tempered with the understanding that, *ceteris paribus*, ethnographic research provides no more nor less solid a basis than quantitative research for policy makers to base their decisions upon (Hammersley, 1992).

3.0 Research methods

The following sections provide the rationale for each of the methods utilized in this research: participant-observation, interviews, and the analysis of documents. The aspects of the review process that each method is proposed to throw light upon are

covered in each section, along with the rationale for the purposive selection criteria utilized with each method. In addition, the practicalities of implementing the particular method are summarized; the chronological utilization of different research methods, which demonstrates a progressively tightening focus upon the substance misuse review and the more senior figures involved in it, is detailed in Table 6. A final section details how ethical issues were addressed in the conduct of this research.

Table 6: Timeline of the research conducted (2006)
Note: I was based principally at the offices of the institution managing the review from April to mid-June; from July to August, I spent whole days at the offices only when attending a meeting. From September onwards, I was present at the offices only when conducting interviews or attending committee meetings.

March Observation of initial planning meetings and public stakeholder meeting for substance misuse review.
April – mid-June 1) Observation of meetings outside of the substance misuse review: a) Public Health and Health Technology committees' deliberations of evidence and drawing up of recommendations. b) Other Public Health team meetings on various matters regarding reviews in progress, e.g. responding to stakeholder comments and the re-drafting of scopes. 2) Participant-observation in substance misuse review managing team (MT) meetings, e.g. responding to stakeholder comments and monitoring and evaluating progress of the review team (RT). 3) Participation in substance misuse review: a) Screening of papers for inclusion in review (n=65) (i.e. acting as first reviewer). b) Checking of other first reviewers' screening decisions (n=73) (i.e. acting as second reviewer). c) Critical appraisal and extraction of data (using screening tool) from included papers (n=38).
July – August 1) Participant-observation in substance misuse review MT meetings to provide feedback on first draft of review and prepare for first committee meeting.
September, October, and January (2007) 1) Observation of substance misuse review committee meetings in which the review was deliberated upon and recommendations drafted.
November 1) Participant-observation in substance misuse review 'fieldwork' meeting (Bristol) where draft recommendations were discussed with practitioners.
September, October, and December 1) Interviews conducted with members of substance misuse review and managing teams.

The review team (RT), who were based at a separate institution from the managing team (MT), were responsible for conducting the systematic review itself (searching databases, critically appraising papers, extracting data, and synthesizing findings in the form of a report). The MT held responsibility for monitoring and co-ordinating this review process within the wider consultative and deliberative processes of the institution.

crucial balance was sought between immersion in, and distance from, the setting and the people within it. Hammersley and Atkinson (1995, p.115) forcibly argue for this "social and intellectual distance" as an essential component of ethnographic research, in the absence of which there is no space in which to conduct the work of critical analysis.

The approach adopted in conducting participant-observation was designed to ease me into the unfamiliar environs of the institution's offices and familiarize me with the rhythms and procedures of working life within them; in short, to establish a place within the setting in which research was to be conducted (Emerson et al., 2001). As such, participant-observation initially included a range of meetings in addition to those taking place for the substance misuse review; these were not intended to provide key sources of data, but were rather intended to familiarize me with the manner in which the institution's review process operated and to stimulate my thinking about the methods used. More prosaically, participant-observation in these meetings allowed me to practice keeping track of the cut and thrust of discourse whilst keeping contemporaneous field notes. In short, I could learn from my mistakes at this earlier stage where the completeness of field notes was not at such a premium.

Participant-observation in the substance misuse team meetings within the managing institution was utilized both to provide contextual understanding of the process and inform the development of interview topics. My involvement in these meetings also played an important social function by allowing working relationships to be developed with the substance misuse RT. In this way, participant-observation was utilized in order to access data that would not be available to a researcher unacquainted with the institution's working practices and the individuals involved (Miller, 1997; Walsh, 1998). There is also a certain understanding of phenomena, such as the idiosyncrasies of ways of working, that can only be gained by experiencing the phenomena in the context of the institution concerned. This understanding included

3.1 Participant-observation

Participant-observation was a key method utilized in this research for its strengths in enabling the researcher to access not just the natural setting in which a SR is conducted, but also the beliefs, interpretations, and perceptions of the people involved with the process of conducting a SR (Brewer, 2000). The method 'works with' the dynamic and contextual nature of naturally occurring settings rather than attempting to impose external control upon it (Boyle, 1994), meaning that the research process is necessarily fluid. Whilst it is advisable in this approach to maintain an awareness of key stages of the research process, problem formulation, data collection and data analysis is an iterative process that cannot be clearly delineated (Walsh, 1998). The primary goal of this approach to research, which could also be said to apply to qualitative research more generally, is to endeavour to illuminate the issues of concern so that they might be better understood (Hammersley and Atkinson, 1995; Schofield, 2000).

The extent to which I, as researcher, was a participant or an observer in the conduct of this research differed according to the individual demands of the research situation. Table 6 catalogues the activities in which I was involved in the conduct of this research. In certain circumstances, such as the committee meetings, I was purely an observer save for some short informal discussions with committee members during coffee breaks. In other situations, such as the appraisal of papers and extraction of data for the review, I was most definitely a participant, albeit one who was also critically observing the process at the same time. Much of the time I spent at the managing institution's offices could genuinely be termed as balancing participation and observation in roughly equal measure; for example, I would contribute to meetings where my involvement in the review meant that I had something useful to contribute, but would observe and take notes when I had nothing to add. Similarly, I would participate in everyday conversation with team members, but would also observe other's social interactions when I was not involved in the conversation. Throughout all of these situations where participation and observation occurred to various degrees, a

work patterns, ways of communicating, organizational hierarchies, acronyms, and non-work related social ties; these enabled me to demonstrate 'insider' knowledge of the institution whilst also maintaining a critical distance. Hammersley and Atkinson (1995) warn that conducting ethnography in culturally-similar settings can lead to difficulties for the researcher in maintaining a critical distance. My experience was to the contrary; despite the ostensible cultural similarities between myself and people who I was researching, I found the research environment to be quite 'foreign' on a number of personal and practical levels (for example, see RD 4587-4624 18/04/06; 4887-4902 28/04/06; 4927-4948 04/05/06).

The ability afforded by the considered practice of participant-observation to operate as an 'insider' of an organization means that vital material that substantively contextualizes the subsequent research methods can be attained. In my use of participant-observation, I have demonstrated a key element of my ontological and epistemological position. Ontologically, my use of participant-observation demonstrates that I do not believe we can know the world adequately without an understanding of the context in which phenomena take place. For example, to conduct an analysis of institutional documents (whether published or internal) would be a dangerous undertaking without an understanding of the context in, and the manner in which, those documents are constructed and utilized. Epistemologically, my use of participant-observation demonstrates that I do not believe that knowledge about the review process can be adequately recalled or articulated by, for example, solely conducting interviews with the team members involved.

This above point regarding my epistemological position also highlights a key reason for undertaking a further aspect of participant-observation in this research; my role as a reviewer on the substance misuse review itself[61]. This role was adopted not simply because I placed value on attaining contextual knowledge, but in order that I

[61] The details of the reviewer work that I completed are contained in Table 6 (see 'April - mid-June', part 3).

could experience the review process, and reflect critically upon it, from the position of a reviewer. Stated in another way, I adopted an epistemological position where some aspects of valid knowledge can only be attained via personal experience. This does not mean that I adopt a 'standpoint' position whereby experience of phenomena equates with valid knowledge (see Mason (2002)). Rather, the experience was intended to inform my critical reflections on the review, although it should also be noted that my broad areas of concern regarding the process of systematically reviewing research were as follows:

- What scope was there for including research findings from lower down the 'hierarchy of evidence'?
- How were issues of context in the research dealt with?
- How were disagreements between reviewers resolved?
- What practical difficulties were encountered in appraising research papers and extracting data from them?
- How were research findings synthesized?
- How was the final document assembled?
- What was the nature of the working relationship between the managing and review teams?
- How was the review process managed?

It should be stressed that this aspect of the research could not involve a sampling strategy, whether purposive or randomized. My participation in the review was on the basis that I would receive papers for appraisal and data extraction in the same way that other reviewers would. Although I was not contractually obliged to 'process' a stated quota of papers, the nature of the CASE studentship meant that I was informally accountable to my institutional supervisor, and as such I felt the need to conduct the work for the review to a high standard and on schedule. In the event, the process was defined by frequent communication; I would inform the RT of any

difficulties I was experiencing and of other research commitments that might delay my reviewing activities. The number of papers that I 'processed' each week was agreed by mutual negotiation, with a slight rise in these numbers towards the end of the appraisal and extraction process in order to assist the RT in meeting the deadline.

Observation of the committee meetings at which the review was deliberated and recommendations drafted required that extensive field notes be made in an effort to do justice to all of the points that were raised and lines of argument pursued. In view of the difficulty of keeping entirely accurate contemporaneous field notes, it would have been ideal to have been able to access the audio recordings and/or the transcripts made of these committee meetings by the managing institution[62]. However, these sources are considered to be highly confidential records of proceedings; obtaining access to them would have been a prolonged process that would require the consent of all committee members, in addition to the senior management team at the institution concerned. It is unlikely that the substantial period of time that this would have required would have been justified in terms of providing substantively richer data for the aims of this particular research. Appendix C contains the research protocol for observing the committee meetings; this was made available to be viewed by the committee members, but in the event there were few who were actually concerned to do so, preferring instead to briefly discuss with me the research in which I was engaged.

Throughout the period of participant-observation a research diary was maintained. The diary consisted of three sections[63]: first, containing the bulk of the field notes and initial analyses, was the section relating to the substance misuse review itself and its associated meetings; second, a 'public' section was kept that contained my reflections upon the process of conducting research by participant-observation.

[62] A professional transcriber produced a verbatim record of the proceedings of each meeting for the purposes of the internal records of the institution.
[63] The three sections were amalgamated during analysis in order to form one document for ease of reference; for this reason, the diary does not follow a strictly chronological sequence, e.g. reflections on the research process appear *after* field notes on the latter stages of the committee meetings.

These two sections, following discussion with some team members who remained unclear about what my research involved[64], were made available on request to any team member who wished to view them[65]. Third, a 'private' section was maintained that allowed me to reflect upon contentious or personal issues relating to the research process. These issues could range from the development of working relationships with individuals to anxieties about my role as a reviewer. The purpose of the 'private' section was not to provide documentary evidence of the research process, but to provide a space in which I could candidly 'think through' issues that substantively impacted upon the research process. The contents of this section (circa 1800 words) were *not* made available to team members, nor are they made available for scrutiny now in the manner in which the 'substance misuse review' and 'public' sections of the diary are.

The ability to reflect upon one's own role in the research process, and how it has influenced the collection and analysis of data, is a crucial aspect of ethnography (Hammersley and Atkinson, 1995; Brewer, 2000). It is therefore incumbent on researchers to develop a critical awareness of how their interactions in the research setting and their necessarily partial and selective framing of questions and events influenced the analysis produced. The research diary is one means by which this critical awareness can be achieved, but it requires a consistent approach to be maintained if it is to allow rigorous reflection upon its contents. In this research, a short form of field notes on meetings and observations of the day-to-day conduct of the substance misuse review were made by hand either concurrently with, or immediately after, they occurred (Emerson et al., 2001; Silverman, 2001). These notes were then typed up, always on the same or subsequent day to the events to which the notes referred, into the relevant section of an electronic version of the research diary. Initial

[64] All team members had been provided (via e-mail, with the additional verbal offer of hard copies if required) with the research protocol and a summarised 'information sheet' (see appendices A and B, respectively) regarding the research, as per the ethical approval given by the UCLH Research Committee 'A' NHS Local Research Ethics Committee.

[65] In the event, no requests were made to view these research diary sections, although regular conversations were held with team members regarding the progress of the research.

analysis of the topics recorded in the field notes was often made in an effort to 'capture' the issues that I perceived to be important in what had occurred and to guide the subsequent re-analysis of the field notes. As advised by Brewer (2000), these initial analyses were clearly distinguished from the field notes in the research diary[66].

3.2 Interviews

My rationale for utilizing interviews as a research method is to generate data that would not otherwise be obtainable. This data can be both a report of 'what occurred' (from the perspective of the interviewee, and subject to the fidelity of their recall and subsequent interpretation), and a means of exploring experiences from an insider ('emic') perspective that substantively impacted upon the production of the substance misuse review. The use of interview data in this way as a resource (about phenomena outside of the interview itself) and topic (through analysis of the language utilized) is not mutually exclusive. Doing so requires a careful path to be plotted that has an appreciation for the insights about events that interviewees may be able to provide whilst not simplistically assuming that such reports are straightforward facts about the world, nor that they are merely social constructions between interviewer and interviewee (Hammersley and Atkinson, 1995; Seale, 1998). In this research, the interviews could also provide contextual detail that allowed assumptions which underlay the production of the SR to be explored.

The above rationale indicate my own ontological assumptions regarding the nature of knowledge, it being possible to locate this knowledge through interviewing people about their experiences and understandings (Mason, 2002). Epistemologically, my position is that, to a degree, the people in this research are able to adequately recall and articulate the elements of their experiences which this research focuses upon. As detailed in the above section on participant-observation, however, my

[66] In this research, analysis recorded in the research diary was differentiated by placing it within parentheses.

analysis of the interviews would be contextualized by the structured reflections upon and analysis of field notes.

The tightly-focused nature of this case study research means that there are only a limited number of people whose experiences and knowledge could be drawn upon. Issues traditionally addressed in the literature with regard to purposive selection criteria, for example ensuring that an adequate spread of respondent characteristics such as age and socio-economic group are covered (see Ritchie et al., 2003) are not applicable in the context of this research. Similarly, there is no tradition of research, in either the construction of evidence-bases, or more broadly in the organizational production of documents, to provide a theoretical basis for purposively selecting interviewees according to certain characteristics. Nevertheless, it remains of great importance to explicitly justify the reasons for selecting the interviewees in this research. The criteria utilized for selecting interviewees was based upon their knowledge of, and participation in, different stages of the review process; these interviewees were members of the team which conducted the review itself (n=2), members of the MT responsible for monitoring the progress of the review and managing the processes associated with it (for example, stakeholder consultation) (n=3), and senior team members who had no direct role in the conduct of the review, but who had methodological expertise and who played a critical role in managing the deliberative process in committee meetings (n=2). Interviewees were selected on the basis of knowledge gained from the participant-observation stage of the research, namely that their role in the production of the review qualified them to speak knowledgably about it.

Following the structure advocated by Mason (2002), preparation for the interviews involved drawing up topic guides consisting of the major and minor research questions. These guides varied slightly in order to provide space for interviewees to speak on their particular area of expertise, but all interviews revolved around the focus

of the methods for conducting the substance misuse review (see Appendices D, E, and F). The interview topic guides were based upon the initial interview topic guides that I devised on the basis of the original proposal for this research and the literature review I conducted during the first year of the research (2005). The topic guides were subsequently revised on the basis of my reflections upon the participant-observation stage of this research and the further reading that I had undertaken with regard to SR methods.

Preparation for conducting the interviews consisted of my developing a well-rehearsed knowledge of how I understood the interview topics to inter-relate and the approximate sequence in which I wished to address them; this mental approach was utilized to provide a structure, yet also allow the interview to follow a conversational course and for me to remain clear on what aspects remained to be covered when this conversational approach resulted in topics being addressed in a different sequence from that envisaged in the interview topic guide (Legard et al., 2003). The semi-structured nature of the interviews allowed for questions that sought clarification or elaboration on particular points; in short, to enter into a dialogue with the interviewee (May, 2001). Necessarily, such dialogue relied upon the cultivation of mutual respect between interviewer and interviewee by demonstrating respect for interviewees' responses, attending closely to those responses and pursuing a line of questioning that demonstrated cognizance of the responses given whilst respectfully probing for further information on issues that remained unclear or unanswered (Legard et al., 2003).

3.3 Documents

The rationale for analysing documents bears certain similarities with that for utilizing interviews as a research method; documents can be utilized not only as a source of information about how an institution works (for example, the methods utilized in the substance misuse review), but also in order to build a picture of the way that an institution accounts for and justifies its actions. It may also be noted that although

documents are indeed inanimate objects, they have the capacity to act back on the wider environment through framing discourse on the subject matter at hand (Bowker and Star, 1999; Prior, 2003; Atkinson and Coffey, 2004). The utilization of documents in this research demonstrates my ontological position (that documents constitute meaningful representations of wider social phenomena) and also my epistemological position (that valid knowledge of these social phenomena may be obtained by a careful critique of these documents).

The documents used in this research can be classified into two groups according to the whether the role that they played in the research was principally with regard to informing the context (i.e. they played a background role) or were subject to explicit analysis (i.e. their role was in the foreground). Two key background documents were the SR methods manual utilized by the institution concerned to direct the conduct of reviews, and the various revisions, and final publication, of the substance misuse review itself. It was vital that I, as a researcher, became highly cognizant of these documents in order to inform my research approach using the other methods, but explicit analyses of these documents' contents was not conducted. 'Foreground' documents included e-mails sent between the substance misuse managing and review team members and, very importantly, the database maintained by the RT in which a record of papers included and excluded from the review (and the reasons for doing so) was kept. Certain research papers, although produced entirely outside of the institution concerned with producing the review, were also considered to be valid documentary sources; their contents could be analysed in the light of the inclusion or exclusion decisions made about them. In considering documents used in this research, it was not a matter of simplistically regarding them as factual records from which the reality of what took place could be deduced, but neither were they treated as phenomena that were entirely socially constructed. In essence, the use of documents required that the tension between their utility in providing a record of how the review proceeded and the

role that the institution played in constructing those documents in a certain way, had simply to be acknowledged and borne in mind when conducting the analysis.

There are strong methodological reasons for utilizing documents in research. All research is conducted within certain resource limitations, so if it is possible to access ready-made sources of data then the researcher can dedicate the saved resources (perhaps most importantly, time) to other areas of the project. Documents are also prime examples of 'low-inference descriptors' (Silverman, 2001); the researcher has not been involved in their production in the manner in which, for example, interview data is produced. Strictly speaking, my role as a reviewer in this research meant that I was involved in the construction of some of the documents used in this research. However, as my contribution was made according to the institution's methods protocol, this could not be considered as an instance of the researcher substantively influencing the production of the documents concerned.

3.4 Ethical considerations

Ethical approval for this research was obtained from committees at both the university at which the studentship was based and, because the research involved National Health Service (NHS) employees, from the relevant NHS Local Research Ethics Committee (LREC):

- University of Plymouth Social Science & Business Faculty Research Ethics Committee approval granted for study (SSB FREC 73/06)
- University College London Hospital NHS Local Research Ethics Committee 'A' approval granted for study (06/Q0505/6)
- NHS Research and Development governance approval was also granted by the institution at which the research took place.

Ethical research practice whilst conducting participant-observation in this research was attained as follows:

- The research methods were agreed with the institutional supervisor.
- The research methods were explained in person to the substance misuse MT during the course of a team meeting (24/03/06), and the opportunity to ask questions was provided. Team members were made aware that they could ask me questions at any stage of the research process.
- All substance misuse managing and review team members and colleagues working on other public health reviews, were provided with electronic copies of the research protocol and information sheet for research participants; hard copies were made available if required.
- The 'substance misuse review' and 'public' sections of the research diary were made available to team members upon request.
- Team members were assured that my taking of field notes was in order to provide material for analytical reflection and the development of interview topics, and not as a source of direct quotes.

I was acutely aware when considering how to address ethical issues whilst undertaking participant-observation that, whilst the structured elements listed above addressed a number of important issues, there remained a substantial element of ongoing negotiation about the conduct of the research with team members. This ongoing element can best be described as a commitment in principle to respect team members' well-being and privacy, as encapsulated in the British Sociological Association's Statement of Ethical Research Practice (British Sociological Association, 2002). This 'grey area', where ethical research practice cannot be assured simply by the researcher adhering to the agreed protocol, is struggled with in the ethnographic literature (see, for example, Brewer (2000)). The tendency is to fall back onto the more

clearly delineated area of the interview in which ethical concerns regarding the right of refusal, anonymity, and confidentiality can be more reliably assured.

In contrast to participant-observation, the discrete nature of the interview (bounded by both the time set aside to conduct the interview and the physically private space in which it is conducted) allows ethical concerns to be more straightforwardly addressed. Consent can be explicitly granted or withheld, and the contrived nature of the interview situation better allows for negotiation between the interviewer and interviewee, without the concern of how others may view what is being discussed.

Ethical research practice was attained in this research whilst conducting interviews as follows:

- Interviews were arranged for a mutually convenient time and location, and took place in a private room.
- Interviewees were provided with an interview topic guide, the interview consent form, the research protocol, and the information sheet for participants at least seven days prior to the interview.
- The consent form (Appendix G) was explained by the interviewer prior to the interview commencing and the opportunity given for questions to be answered. Interviewees could state on the consent form whether or not they wished to maintain their anonymity; consent could also be granted or withheld for the use of direct quotations from the interview in the research. In view of the small number of people involved in the substance misuse review, interviewees were made aware of the possibility of others inferring their identity if consent was granted for the use of direct quotations.
- Interviewees were made aware that interview data (in anonymized form) might be shared with the supervisors named in the research protocol in the course of routine research supervision. Interviewees were also informed that their

comments may inform the development of topics for other interviews, but that they would not be identified in this process.

- Interview transcripts were e-mailed to interviewees within seven days of the interview taking place. Interviewees then had a fourteen day period in which to comment upon, clarify, or withdraw statements from the transcript (or to withhold consent for the transcript to be used in the research) without any judgement being exercised upon their decision by the researcher.
- Data management – *non*-anonymized interview transcripts were kept for the duration of data analysis in order to facilitate analysis. The transcripts, along with the digital recordings of the interviews, were stored on a password-protected laptop, with back-ups stored on a CD and USB memory stick (both also password-protected).

There exists little substantive discussion in the literature with regard to the ethical issues involved in utilizing documents. It might even be argued that this paucity of discussion is evidence of a rare occurrence in the social sciences, namely a research area over which there is little contention; the researcher is analysing an inanimate object that is in the public domain and therefore is not bound by the same ethical concerns that are so important when using other research methods (Abbott et al., 2004). Whilst this argument is tenable to an extent, it does not take account of the fact that to adequately conduct analysis of a document requires the researcher to know about the context in which it was constructed and is utilized; without this knowledge, the analysis risks focusing purely on the text of the document itself, about which inferences regarding its production and utilization may be made that are not adequately grounded in knowledge about the situation concerned (Miller, 1997; Prior, 2003). Obtaining this contextual knowledge, unless the researcher already has extensive

experience in the area concerned[67], can only be obtained through participant-observation and interviews, with all their inherent ethical issues.

Finally, in accordance with the requirements of the body funding this research, the data produced by this research were offered to the *Economic and Social Data Service* (University of Essex) for archiving. Access to this data would allow interested third parties to re-interpret the data utilizing a different framework or to evaluate the validity and rigour of the research by tracing its development (Corti, 2004). However, the dataset was declined for archiving on the basis that there would be limited re-use of the data (*Economic and Social Data Service* 17/09/07, ref. Acq3277/KS).

4.0 Data analysis

Writing about 'data analysis' under a separate section heading risks giving a misleading impression. It might be inferred that the analysis of data took place solely in a distinct phase following the collection of data. Whilst it is certainly true that the most intense phase of data analysis took place after returning from 'the field', data analysis was taking place throughout all stages of the research. At the broadest level, reflecting upon the initial analyses recorded in the research diary provided the basis for drawing up the interview topics and framing the recording of field notes during committee meetings. At a much smaller level, analysis took place in real time during participant-observation and interviews; without this ongoing analysis, the opportunity to critically investigate important points would have been lost. By detailing in this section how data analysis was conducted, I am therefore not claiming that a strict analytical protocol was followed; rather, this section aims to make clear the principles which guided data analysis throughout all stages of the research. In view of the importance claimed in this research for paying attention to the role played by discourses in the methodological development of SRs, the rationale for *not* conducting explicit discourse analysis is also provided.

[67] It might also be added here that it is not experience as such that is of prime importance here, but the process of critical reflection upon that experience.

The overarching principles that guided the analysis of data in this research may be summarized in the terms 'comprehensive', 'iterative', and 'reflexive'. Data treatment was comprehensive, not in the sense of having researched an entire population, but in terms of ensuring that the development of analyses took account of all the relevant data (Silverman, 2005). Taking this approach also stimulated the development of new analyses, for if an explanation of data was developed for which there were deviant cases, was this because the explanation was inadequate or because the deviant case was an example of phenomena not previously understood? It was necessary to exercise care using this approach, for the risk is that the researcher could begin to explain away deviant cases on the basis that they constituted examples of other phenomena rather than the developing theory being wrong. This is where iteration in the analytical process becomes so important, as the researcher treks back-and-forth through the analytical categories by which the data was coded[68] in order to test out the explanations that they are developing (Morse and Richards, 2002; Spencer et al., 2003; Silverman, 2005). It is necessary for the researcher to exercise "analytic nerve" (Hammersley and Atkinson, 1995, p.210) when using this approach if the development of explanations is to be rigorous and the temptation to reach analytic conclusions that simply accord with the researcher's prior conceptions is to be avoided (Hammersley and Atkinson, 1995). The approach also requires a commitment to stick at the task when a carefully developed hypothesis is shown upon further analysis to be inadequate.

The reflexive element of data analysis is with respect to the role played by paying close attention to the impact upon the production of data that I had in my role as researcher. As with all elements of the research process, practising reflexivity effectively requires that a balance is struck. The researcher needs to maintain an awareness of how they have influenced the production of data, but without letting this

[68] A qualitative data analysis package (NVIVO 2.0) was utilized to store and code the research diary and interview transcripts according to the thematic categories developed; the software package facilitates the retrieval of all instances of a particular thematic category across all of the data sources, and allows analytical categories to be rapidly re-configured to enable different perspectives to be gained on the data.

awareness dominate the analysis (Silverman, 2001). In view of the fact that the researcher does not play a static role throughout the conduct of the research, an awareness of how that role changes over time and in different situations with team members is argued to be crucial for adequately contextualizing the analysis of data. In this research, my role as researcher changed over the course of time from one of neophyte (needing introduction to work patterns and policies), to reviewer and researcher (needing monitoring for adherence to the review methods protocol and for the use of sensitive information relating to the review), and finally to a critical participant who was able to engage in discussion about the development of SR methods (whilst also working within the agreed terms of the CASE studentship).

The claim is made in this research that attending to the role played by discourse in the discussion and development of both SR methods and conceptualizations of substance users is of great importance. This claim is based on the argument that an awareness of the ways in which discourse can permit the consideration of certain options and foreclose others is essential for attaining an adequate understanding of the SR and policy making process. However, no explicit discourse analysis is conducted in this research; discourse is simply treated as another element within the ethnography of the substance misuse review. This requires some justification, for there exists a substantial literature upon methods of discourse analysis that is rooted in the work of Michel Foucault (principally Foucault, 2002) and which has been substantively developed over time (Potter and Wetherell, 1987; Fairclough, 1992; Wetherell et al., 2001). First, it is in the nature of this research to draw upon a wide range of sources, and to focus in upon analysing documents and interviews solely in terms of discourse would severely limit the time available to focus the ethnographic lens in wider terms. This is not to claim that a purely discourse analytic study of the phenomena in this research would not be justified; it is simply that it is not the route chosen for the purposes of *this* research, which seeks to gain a wider understanding of the SR process. Second, in view of the labour-intensive nature of conducting discourse

analysis, there is the pragmatic issue of devising a purposive sampling strategy for the copious quantities of text contained in documents and the transcripts of interviews and meetings. The analysis of SR processes is an undeveloped 'science', and I am personally not confident that a rigorous and defensible strategy could be implemented in this respect. Third, and finally, there is the issue of data access. The majority of the interviewees in this research, presumably because of the sensitive nature of the interview topics, did not give consent for direct quotations from the interviews to be used in this research. In a similar manner, the transcripts made by the institution for the purposes of recording the proceedings of committee meetings, were not available for research purposes. The use of discourse analysis would therefore have been restricted, for the most part, to publicly available documents that would have provided only a partial (and arguably insufficient) source for analysis.

Summary

The rationale for adopting a case study approach, and for emphasizing construct validity in the conduct of this research, have been clearly set out. My ontological position as a realist has been elucidated so as to facilitate both understanding and critique of the research conducted. The research methods utilized (participant-observation, interviews, and documents) have also been described and justified, in addition to the ways in which an ethical approach to the research was attained. Finally, the analytical approach taken to the data, and the rationale for making careful use of discourse analysis, have both been explained.

Chapter 5

Taking research apart: Appraising papers and extracting data

Conducting a systematic review (SR) requires evidence to be extracted from those sources which are considered to be of sufficiently high quality. This chapter thus provides an analysis of this initial process using a case study of a SR of effective interventions for preventing or reducing substance misuse by vulnerable and disadvantaged youth. Chapter 6 provides an analysis of the process by which the extracted data was synthesized into the form of evidence statements that were subsequently utilized by a committee in order to write policy recommendations. Chapter 7 considers how the critiques of the SR in this case study can be used as a means of taking forward the development of SR methods for social policy.

In this chapter, a summary of the key stages of the review and the issues arising at each of these stages is first given in order to provide an overview of the review process and to clarify the roles of the managing team (MT) (responsible for overseeing and co-ordinating the review and consultative process) and the review team[69] (RT) (responsible for searching for, screening and appraising studies, and extracting data for synthesis in the review). The chapter then progresses to an analysis of how the review proceeded to take research apart, both methodologically and in terms of extracting data for use in the review. The benefit of hindsight is used to consider the rationale for excluding certain studies from the review, and a purposive sample of excluded studies are further examined for the contribution that they might have been able to make to answering the review's questions. An analysis of the practicalities of conducting the SR is also presented in order to facilitate an understanding of the production of the review within an institutional context. The analyses draw upon interviews conducted with members of the managing and review teams, email exchanges between these team members, and the research diary that

[69] In this case study, the RT was a team external to the managing institution.

was kept as part of the process of participant-observation in the SR. The analysis of excluded studies utilized the RT's database (which recorded all study inclusion and exclusion decisions, and the rationale for them) as a means of informing the purposive sample of studies for further analysis.

1.0 Chronology of the review

The topic for the review was outlined, with reference to current policy objectives in England[70], by senior staff within the MT during the second half of 2005. This planning sought to build upon the evidence-base of two previous reviews that had identified the shortcomings of universal educational approaches to reducing substance misuse in the young and the need to better understand what was effective in preventing or reducing substance misuse in at-risk groups[71] (Canning et al., 2004; Edmonds et al., 2005). In November 2005 it was confirmed by the relevant government department that the topic addressed the appropriate issues; subsequently, a formal referral was issued to the institution at which the MT were based ('the institution') to complete the work. In December 2005 the review was publicly announced; the draft scope of the review (stating in greater detail the rationale for the review, the population, interventions, comparators, and outcomes to be covered) was posted on the institution's website in February 2006[72] and interested parties were invited to register and comment upon it. On March 10th, a meeting was held with stakeholders[73] at the institution's offices to publicly discuss the scope. In addition, it was necessary for stakeholders to formally submit their comments to the institution in the subsequent fortnight. These submissions were collated and responded to in a public document posted on the institution's website following an internal MT meeting to discuss the

[70] The policy recommendations that were to be informed by the review were to be applied only to England, not Wales, Scotland, or Northern Ireland.
[71] At-risk groups were defined in the review's scope as young people who; had family members who misused drugs, were offenders, had behavioural conduct disorders or mental health problems, had been in institutional or foster care, were homeless, were socially excluded (for example, some black and minority ethnic communities and some socio-economically deprived groups), were excluded from school or who were persistently truant.
[72] All subsequent dates referred to in the review process are in the year 2006, unless otherwise stated.
[73] For the institution, a stakeholder is considered to be a member of an organization that represents the interests of people who will be affected by the phenomena examined in the SR. Organizations are required to register their interest in particular review topics with the institution in advance; they are also required to state who they consider themselves to represent and why they have an interest in the topic in question.

comments (March 24th), a teleconference with the RT (March 29th), and feedback from the internal group concerned with overseeing the development of guidelines.

Beginning in early April, the RT developed a search strategy for the review in consultation with information specialists at the institution. This search returned 16,621 'hits', an unmanageable quantity for full screening. The RT rapidly screened the titles of these papers to leave a total of 4507 possible papers for inclusion; this number was reduced by a further rapid (non-formalized and undocumented) screen to give a total of 880 papers for retrieval for the purposes of a full screen (that is, a more detailed examination of whether the papers met the criteria for inclusion in the review). The full screening of these papers took place from late April to early May; the first reviewer's decision upon inclusion or exclusion was checked by a second reviewer. At a teleconference with the RT on May 8th it was stressed that the RT should work closely with the MT team with regard to managing the intense workload; a further decision was taken to give a further month from the day of the teleconference for inter-library loan papers to arrive[74].

The 222 papers that met the criteria for inclusion in the review were critically appraised (that is, rated on the hierarchy of evidence) in the period from May 8th to June 9th. Outcome data was also extracted from the included papers in this time period. Liaison between the RT and MT during this time took the form of email exchanges on specific points and a weekly update of progress by phone, usually on a Friday afternoon. The submission of an interim report on the review to the MT brought forth concerns that the RT had an inadequate structure for the write-up of the review and that the sheer volume of studies being appraised was inhibiting this planning process. Concerns were also voiced in a teleconference (June 6th) that qualitative research was barely included in the papers being appraised; the RT stated that this situation was

[74] The eventual outcome was that 71 papers identified for screening were not obtained; 24 of these were not obtainable from the British Library in the time available and 47 were non-English language papers.

simply a result of using the methods manual protocol that was oriented towards appraising quantitative studies. The MT also began meeting with the institution's public involvement unit during this time period in order to develop a plan for the inclusion of non-experts in the review process[75].

2.0 Searching for, screening and critically appraising papers, and extracting the data

The process of searching for, screening and critically appraising papers for inclusion in the review was governed by a formal protocol (National Institute for Health and Clinical Excellence, 2006b, the 'methods manual'); in particular, this manual details the quality criteria upon which different types of studies should be assessed. The methods manual is utilized in conjunction with the scope (a tightly worded outline of what the review will address and the policy context in which it is set) in order to direct the review process in an effort to attain rigour and transparency. This section draws on the research diary and interviews with review team members to provide an analysis of the difficulties that were encountered in following the methods manual. Crucially, this analysis will consider how discourses about SRs used by the managing and review teams highlight the way in which methodological problems are conceptualized, and thus how suggested solutions to those problems are defined.

2.1 Taking research apart: A technical procedure?

Within the institution it is considered to be important to adhere to the methods manual. In the substance misuse review, the RT codified the protocol in the form of a Microsoft Access database onto which all of the data and decisions were entered in the format that was detailed in the methods manual. Analysis of the research diary and of the interviews conducted with review and managing team staff does, however, point to a number of areas where difficulties occurred and where the RT made strenuous efforts to adhere to the methods even when to do so made life difficult. It is important to

[75] The inclusion of two members of the public (who have personal experience of the issues upon which the review focuses) at committee meetings is standard practice within the institution as a means of representing a lay perspective on the issues discussed (National Institute for Health and Clinical Excellence (2006a)).

note here that RT members were not blindly following this technical protocol, but were doing so because they genuinely believed that it would result in the production of the most rigorous evidence-base. Despite the difficulties that they encountered using this protocol, they were still able to produce a rational account of why they considered it to be the best (if not yet ideal) method of reviewing evidence.

In considering the responses of RT members with regard to the process of conducting the review, it is important to emphasize both the commonalities and the differences between them. Whilst it may be argued that there is an overarching 'adherence' to the traditional model of conducting SRs (see Chapter 2, Figure 2), it is equally important that the ambivalent nature of some responses given are explored for the insight that they can give into the process of conducting a review. However, it should also be acknowledged that review team members maintain a strong 'allegiance' to the process of systematically reviewing evidence; solutions to problems encountered are framed in the sense of how to improve upon the current methods rather than a more fundamental revision of those methods.

The 'strong adherence' to the traditional SR methods was evident in all of the responses given at interview by the RT, but was notably weaker in some of the responses from more senior staff members of the MT. The MT tended to justify the review's procedures on the basis that they were the best tools for the job, whilst the RT highlighted a number of areas where the full use of appropriate SR techniques had been limited by the review process. To expand upon this: a member of the MT stressed the comprehensive nature of the review, and most importantly, how it was working *from the evidence* – if there was evidence for the effectiveness of abstinence-based approaches to substance misuse, then the review would have picked up on this (ID 1 53-60). This belief in the strong utility of the SR processes contained in the methods manual is further expanded upon by the same interviewee in justifying the prior identification of populations and interventions of interest for the scope (based upon the

initial referral) as a means of ensuring rigour and reducing bias in the conduct of the review (ID1 86-90), and again by a more senior member of the team with regard to adhering closely to the scope throughout the review process (ID2 148-156).

The RT, whilst anxious to stress that the overall rigour of the review was not adversely affected, pointed out a number of areas where they felt that the timeframe for completing the review had impacted strongly upon it. The RT held the view that the large volume of literature which the search terms had yielded and the time available in which to review it, had limited their ability to conduct the review totally in the manner which they thought appropriate for a 'systematic' review. For example, at the screening stage, the titles of papers and abstracts (if available) were read in order to determine whether they were appropriate for inclusion in the SR. In the substance misuse review, this was conducted by a sole reviewer (ID3 355-362). Standard practice within SRs is to utilize two reviewers in order that disagreements may be discussed and the likelihood of mistakes reduced. In a similar vein, time limitations meant that there was no opportunity for an additional handsearch of journals to check for other relevant research that the database search may have missed, nor to follow-up on the citations provided in papers or reviews (both systematic and non-systematic) that had met the review's inclusion criteria (ID3 375-394).

However, in the interviews conducted for this research, the MT did not dogmatically defend established SR methods, nor the methods manual; faced with difficulties in the review process, or challenged about its coherency, they would thoughtfully propose improvements intended to improve the rigorous utilization of evidence. These ideas were notable for the way in which they met the underlying philosophy of the traditional method of conducting SRs. For example, with regard to the exclusion from the substance misuse review of non-systematic reviews, it was proposed that these could legitimately be brought into the review, but that a separate, fully worked up search strategy would be required in order to ensure a comprehensive

treatment of that area (ID5 455-466; 485-492). Furthermore, the non-systematic reviews would need to be rated according to criteria similar to those which SRs were (ID2 306-316). Similarly, with regard to the framing of the review question in terms of *effectiveness* (which, given the hierarchy of evidence utilized in the methods manual would, *ceteris paribus*, prioritize randomized-controlled trials (RCTs)), it was proposed that a slightly modified scope that clearly stated the importance of evidence regarding implementation issues would have resulted in the inclusion of a range of qualitative studies (ID2 399-410).

As would be expected when utilizing the traditional approach to SRs, where the hierarchy of evidence clearly sets out a gradation (in terms of establishing the effectiveness of an intervention) from rigorously conducted RCTs (strong evidence) down to expert opinion (weak evidence), a major concern of the RT was to ensure that studies were correctly categorized according to study type. Another way of stating this is that the strongest emphasis was put upon the internal validity of studies as a means of assessing their utility as strong forms of evidence, rather than a consideration of their external or construct validity. One way in which this was manifested in the conduct of the review was that a not inconsiderable period of time was spent by the managing and review teams debating what actually constituted a particular study type. In particular, differentiating between non-randomized controlled trials and controlled before and after studies proved contentious (RD 1755-1757 06/07/06; RD 1880-1903 10/07/06) and the subsequent checks regarding the appropriate classification of studies according to the revised criteria proved time-consuming (ID4 483-496).

2.2 'What works' and the knowledge needed to implement it

The emphasis placed in the review upon internal validity demonstrates the goal when conducting an SR to establish 'what works'; if the best way to assess whether an intervention is systematically related to an outcome is to utilize an experimental study design, *ergo* the randomized, controlled form of this design will produce the most

rigorous evidence. This thinking was evident in many of the interviewees' thoughts upon the review process. It took the form of a recognition that whilst such reasoning could lead to some forms of evidence being overlooked, it was essential for rigorously establishing what was effective; moreover, the evidence that was initially 'missed' through utilizing this approach could, through a widening of the study inclusion criteria, be utilized to inform processes such as implementation of the effective interventions (ID1 145-151 and 156-179; ID2 394-407 and 437-445; ID3 242-250 and 515-520).

The desire of members of the RT to stress that they had a broad-minded and inclusive approach to evidence illustrates the importance that they placed upon seeking legitimacy for the review's results amongst a wide audience. Arguably, they were at pains to justify both the rigour of the review in terms of the prevailing institutional review methods *and* to stress how the review had gone beyond this methodology in order to address the particular problems posed by the topic. Where the review methodology was acknowledged to be insufficient for the particular demands of the review, it was proposed that refinements to the methodology would address the problems identified.

Whilst it is not the case that these different approaches are incommensurable, the way that they are currently understood led to RT members having to either fall back upon the dominant methods within the institution or engage in quite heroic efforts to reconcile the different approaches. On occasion this may occur because of the conflicting loyalties felt by the RT between adhering to the methods manual and adopting a more inclusive approach to evidence that they feel is advocated within the wider Evidence-Based Policy and Practice (EBP) field. For example, the MT team questioned the preponderance of RCTs that the RT appraised as suitable for inclusion in the review (RD 1777-1780 06/07/06), to which the RT responded that this emphasis was inherent to a utilization of the review methods in the manual (RD 1890-1903 10/07/06; see also ID3 498-513). Discussion hinged upon whether a *non*-randomized controlled trial was necessarily less rigorous in the context of the substance misuse

review given the many difficulties of rigorously operationalizing a fully randomized, controlled trial in this field. No real consensus was reached, and the decision was made to adhere to the study assessment criteria (prioritizing randomization) in the methods manual, with the *caveat* that this should be noted in the review as a limitation.

The issue of the 'exclusion' of qualitative research from the review provides a strong example of how the two discourses of (a certain form of) scientific rigour (embodied in the methods manual) and one of wide inclusiveness of different forms of research are irreconcilable in the forms in which they are currently understood by the managing and review teams. As detailed above, on a straightforward level, some interviewees recognized that utilizing a hierarchy of evidence that places RCTs as producing the most rigorous forms of evidence will most likely produce an 'evidence-base' consisting of a preponderance of RCTs. This outcome is rationalized on the basis that study quality is independent of study type; potentially, a judgement could be made that a highly rigorous controlled before and after study is a higher form of evidence than a poorly conducted RCT. It is possible to make such a judgement within the bounds of the dominant discourse about scientific validity within the institution. However, difficulties occur where qualitative research is concerned; given that within the dominant discourse it is considered a weak form of evidence (for establishing the effectiveness of an intervention), it is not difficult to understand why improved review methods are proposed that consist of separate reviews for effectiveness (prioritizing RCTs) and for an understanding of implementation and acceptability issues (prioritizing qualitative research).

However, the neat demarcation into 'effectiveness' and 'implementation' issues does not address the deeper issues that permeate the SR process. This can be illustrated by the ambivalent attitude taken towards the role of experts in the review process. On the one hand, experts are valued for the rich understanding of the field that they can bring to the review process, but on the other, the review process relies to

a large extent upon the impartiality of the reviewers and their adherence to a technical procedure in order to minimize bias within the review. Whilst it is reasonable to make a distinction between accepting what an expert states simply because of their status, and a process of deliberation that is informed by the interpretations of experts, this does not address the deeper issues of the exclusion of non-systematic reviews (conducted by experts) from the review. Neither does it address the difficulty of separating knowledge from expertise. Arguably, there is not a neatly bounded stock of knowledge waiting to be collated using the SR process. Rather, it is inherently tied up with expertise, but the dominant discourse surrounding SRs cannot account for rigorous knowledge produced in this way. It is perhaps unsurprising that utilizing this dominant discourse in an attempt to justify the selective inclusion of 'expert knowledge' in the SR process resulted in members of the MT struggling to make statements that were not highly ambivalent, or even contradictory (ID1 213-227; ID2 350-369 and 385-399; ID5 150-223).

Although a number of difficulties were identified by the managing and review teams with regard to following the methods manual, at other times this adherence was turned into a virtue that excused the non-consideration of wider issues surrounding the SR process. Most pertinently, the nature of the review process being part of a wider consultative process (in which stakeholders were invited to comment upon the scope of the review, its findings, and the proposed guidance) was used as justification for not, for example, considering the appropriateness of revising the scope in the light of what the search strategy had unearthed[76]. Although the idea that revising the scope might be useful was not dismissed by the MT, the impracticality of repeating the consultation with stakeholders in order to validate it was argued to be impractical, particularly with regard to the tight timeframe in which the review had to be completed (ID1 559-566; ID2 175-180; ID5 295-315). Notably, the RT did not share this commitment to

[76] Note that the revision of a scope is usually considered, in the traditional approach to SRs, to be indicative of bias.

consultation with a broad range of stakeholders, instead emphasizing the importance of building the scope and conducting the review in response to the research findings that an initial literature search, and a subsequent systematic database search, uncovered (ID3 254-276). This analysis is not to suggest that the RT did not consider stakeholders' views important, but rather highlights the difficulties that the RT felt they faced in conducting a review that did not build upon (in their view) the structure and understanding which the research field had of the problems that were the subject of the review.

Although both the managing and review teams detailed the manner in which scientific rigour was the *sine qua non* of the review process, this did not preclude them from identifying areas where elements of practice that were not completely transparent crept into the review procedures. This could occur at a number of stages; the judgements that were made as to whether a study really did meet the inclusion criteria or not (ID5 422-430), the prioritization of some outcomes over others when extracting data from a study with a large number of recorded outcomes (ID1 456-466), and the background knowledge that an expert reviewer utilizes in order to inform their decisions about study inclusion/exclusion and data extraction (ID2 641-656). These small examples of subjectivity were not considered sufficiently serious to damage the rigour of the review process as they were viewed as inherent to the process. Furthermore, members of both the managing and review teams considered that any negative impact could be ameliorated by a rigorous management process and the transparency that could be attained by the documentation of the review process.

3.0 The exclusion of papers: What evidence was excluded from the review and why?

The screening of study titles and/or abstracts resulted in the exclusion of 15,741 papers from the review on the basis that their content was not relevant to the criteria set out in the review's scope. Screening of full papers (n=880) resulted in a further 587

papers being excluded; it is these papers that are the subject of the analysis in this section. It is worthwhile further investigating these excluded papers for the insight which may be obtained into what knowledge was (systematically) excluded from the review and for the different, potentially valuable for the making of evidence-based policy, directions that the inclusion of such knowledge could have taken the review and the subsequent deliberations of the committee in. It is also worth noting at this juncture that it is the documentation of decision-making in a SR which makes this analysis and critique of the review possible. However, despite this transparency, it cannot be claimed that it is a straightforward process to critique a SR. The process of analysing the database containing the decisions made, developing a thematic schema for the reasons why studies were excluded, and obtaining and analysing the papers that were excluded is a resource-intensive process. Moreover, this process is one that is unlikely to be feasible for those not conducting research on the process of a particular SR, as I am doing here.

The large number of excluded papers necessitated the development of a purposive sampling strategy so as to allow a focus of sufficient depth on certain themes within the excluded studies. A randomized sample would have been unjustified given the diversity of the papers (even when divided into sub-categories) and the small numbers contained in some of these sub-categories. In an effort to strike a balance between depth and breadth in the analysis of the excluded studies, the papers were categorized as follows (see Appendix H for citation details):

1) Papers (n=352) where the topic under investigation clearly fell outside of the parameters specified in the review's scope - for example, where a universal intervention was evaluated, where the study subjects were above 25 years of age, or where there was no sub-analysis of outcomes relating to substance misuse (distinct from the use of alcohol and/ or tobacco).

2) Papers (n=3) that were concerned with the costing or modelling of the impacts of substance misuse.

3) Papers (n=93) where the effectiveness of an intervention was not assessed to have been adequately established – for example, where no explicit intervention was evaluated, or where outcomes were judged to have been insufficiently reported.

4) Papers (n=139) that reviewed and synthesized papers pertaining to the review's scope, but which did not meet the criteria of a SR.

The further analysis of excluded papers presented here is limited to those papers categorized as (4). Category (1) papers are not analysed, as their focus lay outside of the review's scope; category (2) papers were only excluded in the sense that they did not inform the 'effectiveness' review – they were, however, forwarded by the RT to the external team responsible for the cost-effectiveness review that was to later inform the deliberations of the committee; category (3) papers potentially offer a fertile route for further exploration and analysis, but time-limitations preclude their analysis here. Category (4) papers are contended here to comprise the most informative group of papers for further analysis in view of the potentially wide span of knowledge that they may have been able to offer the review.

In order to inform the purposive sampling of the category (4) papers, abstracts were obtained for 101 of the 139 papers[77] in order that a better understanding of their focus could be attained, and sub-categorization of these papers performed. The sub-categorization of these papers is listed in Appendix I; again, in order to analyse these excluded papers adequately, a tighter focus on certain sub-categories has been necessary, and thus only the papers in three of these sub-categories are further investigated here. These papers are in the sub-categories of papers focusing on black and minority ethnic groups, upstream interventions (i.e. those not focused upon the

[77] Twenty-two of the citations classified as non-systematic reviews were books or book chapters and thus did not have abstracts available; a further sixteen abstracts were not available within a reasonable timescale – this was defined as not being available from major database sources (Web of Science, CINAHL, IBSS, and PsycINFO).

individual), and papers offering different conceptualizations of the problem of substance misuse and how it may best be addressed. Utilizing the benefit of hindsight, these three sub-categories were purposively selected for their potential to usefully inform some of the areas which the committee had the most difficulty with in producing evidence-based policy.

3.1 Excluded papers relating to interventions with black and minority ethnic groups

Each of the four excluded papers relating to interventions with black and minority ethnic groups (Cervantes and Pena, 1998; Yuen and Nakano-Matsumoto, 1998; Kumpfer et al., 2002; Hawkins et al., 2004) drew upon research conducted in the United States of America (US); the importance of cultural awareness in the delivery of interventions designed to prevent or reduce substance misuse was a common conclusion, as was the need for cultural-specificity that consisted of more than simply the use of research participants' mother tongue or the involvement of researchers from the ethnic groups concerned. Further common themes were the dearth of rigorous research in this area, the problems created by the assumption that ethnic groups constitute homogeneous wholes (when factors such as level of acculturation, or whether the person is a recent migrant or second or third generation US citizen may be of greater significance), and the tension between developing programmes that acknowledged this heterogeneity whilst also acknowledging that important commonalities exist between (for example) Asian-Americans, African-Americans, and Native American Indians. In short, the papers were questioning the validity of commonly-used constructs in the research, but stopped short of proposing more refined constructs that could frame future research.

The limited amount of evaluation specific to black and minority ethnic groups which these reviews were able to draw upon limited the scope of two of the papers (Cervantes and Pena, 1998; Yuen and Nakano-Matsumoto, 1998) to a summary of what are generally accepted to be the key issues in developing culturally-appropriate

interventions rather than a synthesis of findings from research designed to test these interventions. In this way, their potential for contributing to the substance misuse review was limited. Kumpfer et al.'s paper (2002) offered a brief overview of research evaluating the effectiveness of 'culturally-specific' versions of the Strengthening Families Program; however, as the database search conducted for the substance misuse review had identified a number of other (far more detailed) papers evaluating this intervention, the further contribution that Kumpfer et al.'s paper could make was limited.

Hawkins et al.'s paper (2004) provides a quite cautious approach to making inferences from the results contained in the papers reviewed with regard to the American Indian population, stressing that the heterogeneity of this population makes generalizing from even one segment of this population to another problematic. It is quite legitimate that the substance misuse review should (in the absence of details regarding factors that might make the findings applicable to black and minority ethnic groups in England) not draw upon these findings, but doing so does leave unanswered the crucial question of how the external validity of *any* of the studies in the substance misuse review was to be rigorously assessed. However, Hawkins et al.'s paper does introduce a framework (an interim evaluation of a programme using the approach having been made at the time of the publication of the paper) that could be utilized to structure the analysis of other research upon 'culturally-specific' interventions in the substance misuse review. This framework emphasizes the integration of interventions (such as coping skills training) developed within the mainstream culture with the elements of culture (such as myths or songs) that are particularly meaningful to the people whom the intervention is aimed at. Arguably, this represents an alternative way of framing relevant constructs which could have informed the search, appraisal and synthesis strategies in the case study SR. Arguably, it could also have facilitated the interrogation of research findings in such a way as to produce a more usable review to inform the deliberations of the committee.

3.2 Excluded papers relating to upstream interventions

The excluded papers relating to upstream interventions do not synthesize evidence in the manner required by the review; studies are not assessed for their quality and effect sizes are not formally cumulated, relying instead upon a narrative summary of the outcomes. The programmes described in two of the papers (Greenwood, 1992; Schinke et al., 1997) provide few details regarding the research methods used, thus making assessment of the validity of the conclusions reached problematic. Schinke et al. (1997), whilst providing great detail regarding the processes by which the programme described is proposed to increase the coping skills of high-risk youth (and thereby decreasing their propensity to misuse substances), provides no formal evaluation of the programme concerned. However, both of these papers highlight a substantive problem within the field of research upon the prevention of substance misuse that may have usefully informed the review. This knowledge is not about *what* interventions are 'effective', but the inherent difficulties of obtaining this knowledge. Thus, the difficulty of recruiting and retaining high-risk youth in rigorously evaluated programmes is highlighted in both papers as a key issue that holds back knowledge on the subject, but the proposed solutions are quite different. Greenwood (1992) advocates the greater use of experimental study designs, whilst Schinke et al. (1997) views such designs as inherently unsuitable for evaluations in this area in view of their need for "stable and compliant" (p.52) research subjects. Arguably, this debate about the relative importance of internal validity, and whether or not it can realistically be attained in research upon high-risk youth, should have informed the initial conceptualization of the review, as well as the deliberations of the committee.

Other excluded papers in this category highlight aspects of the review and its deliberation by a committee that might have usefully followed a different path. Ruffolo et al. (2003), in reviewing evidence for the effectiveness of programmes for pre-school aged children (as latterly implemented in England in the form of 'Sure Start'), conclude that the outcomes are mixed. Perhaps unsurprisingly, programmes which were

possible to implement with great fidelity produced better outcomes, but the reality of implementing such programmes in the context of the everyday delivery of social services meant that outcomes were not so positive. Moreover, Ruffolo et al. (2003) argue that the deleterious impact of the wider environment could create problems for children growing up that would far outweigh the positive effects of the early intervention programme. If the implications of this argument were to be taken on board for the review, the conduct of the review would require substantial revision. First, if programmes are to be implemented where resources are limited, it is not the cost-effectiveness of the programme in a (near) ideal environment that needs to be assessed, rather its cost-effectiveness in the imperfect environment of everyday services[78]. Second, if it is the wider environment that is largely responsible for inhibiting the effectiveness of the early intervention programmes, then the politically unpalatable solution would be to address these wider determinants rather than focus upon interventions that target the individual. However, the hierarchy of evidence utilized in the review persistently drives the inclusion of evidence that is considered to be rigorous; and this evidence is only that which is measurable as an outcome from discrete packages of interventions directed at an individual and/or their family, rather than the broader determinants of health highlighted by Ruffolo et al. (2003). Implicit in this formulation is a construct of substance misuse that is centred upon the behaviour of the individual as opposed to (for example) an understanding of the individual in the context of the effects that the environment has upon their substance misuse.

The review of papers provided by Schaps and Solomon (2003) also provides evidence that 'interventions' at a broader level than the individual may be of substantive importance in this field. Although the authors are clear that their conclusions are based upon a relatively small body of evidence (consisting of both correlational and intervention studies), they conclude that preventing substance misuse cannot be viewed as an educational component that is delivered in isolation; rather, it is

[78] Stated in Campbell's terms, the external validity of the study would be the priority here.

the whole school environment (broadly stated, the quality or otherwise of teacher-student relationships and the degree to which students feel included and valued in the school) that impacts upon a range of behaviours of which the problematic use of substances is but a part.

3.3 Excluded papers outlining different conceptualizations of substance misuse in the young

The purpose of further investigating excluded papers that provide different conceptualizations (or 'constructs') of substance misuse by young people is not to suggest that these approaches offer a panacea, but to make clear that the review itself represents *one* approach that could have been taken towards addressing substance misuse amongst high-risk youth. Arguably, the approach taken in the review fits with Cuijpers' (2003) categorization of approaches to substance misuse prevention programmes, where, since the 1980s, the social influence model (where the individual's ability to 'resist' the use of substances) has been dominant. The issue is not so much to do with the review adopting an approach that utilizes the social influence model, as with the presentation of the review and the subsequent recommendations as objective knowledge that sits apart from the social and political milieux which created it. What directions could the review have taken and what knowledge would have been produced if one or more of the approaches outlined in the excluded papers in this section (O'Connor and Saunders, 1992; Blackman, 1996; Lilja et al., 2003) had been further utilized in order to investigate the state of knowledge?

O'Connor and Saunders (1992) detail a number of problems that they consider to hold back the development of effective drug education, of which two will be outlined here. First, it is argued that the portrayal of substance use (and hence the moral rightness of resisting using them) as uniformly deleterious conflicts sharply with children's experiences where they observe others' substance use. This substance use is not seen as something negative; O'Connor and Saunders (1992) posit that, amongst other things, it is seen as exciting, pleasurable, risky, and naughty. Unless the use of

substances can be understood in the context of these motivations, it is argued that it is unlikely that any long-term impact upon the safe use of substances amongst youth will be achieved. Second, O'Connor and Saunders (1992) contend that models of health behaviour that attribute a detached rationality to the decisions that are made with regard to drug use are untenable. These decisions are made within a dynamic cultural environment, and may be ambivalent and/or emotional rather than calculating. Moreover, such a conception seriously questions the weight that can be given to research that records 'intention to use drugs' as an indicator of the effectiveness of an intervention.

Lilja et al. (2003) detail eight different models that they argue can contribute to an understanding of the processes that lie behind an individual's use of substances. It is stressed that none of the models provide perfect understanding; rather, they throw light upon different aspects of the phenomenon, their utility being in their complementary rather than exclusive use. For example, cultural models highlight the manner in which youth identity (through membership of sub-cultures) is actively constructed by the individual rather than passively absorbed, whilst control models highlight the impact that an individual's perceived locus of control can have upon their behaviour in any particular situation. Lilja et al. (2003) argue that the complementary understandings that these models can provide are vital if an adequate understanding of the complex interplay of factors leading to substance use is to be attained, and effective interventions developed to address these issues. In particular, the role of substance use in facilitating contact with others and the development of sexual relationships is argued to be a key area where understanding needs to be developed.

Blackman (1996) takes issue with the dominant conceptualization of young people's substance use as 'deviant', arguing that this portrayal of young substance misusers as 'outside the norm' impoverishes understanding of the phenomenon. For example, the understanding of substance misusers as victims of 'drug pushers'

(whether peers or adults) does not facilitate the development of a richer understanding of the social processes leading to an individual's drug use. Blackman (1996) also criticizes the tendency to homogenize young people's substance use, whether through treating the use of different types of drugs in the same way or assuming that there is a pathway from cannabis to heroin use and subsequently death. Finally, Blackman directs his focus strongly upstream in arguing that consumer capitalism should receive critical attention for the manner in which it contributes to the phenomenon of substance misuse in the young. It is argued that the dominant cultural values of consumer capitalism (individuality, choice, the rights of the customer) and the wide promotion of purchasing as a means to attain immediate gratification have impacted substantively upon the lives of young people. In this view, the rise in substance use amongst the young is viewed as inextricably linked to changes in the wider political-economic environment.

4.0 The practicalities of conducting a systematic review

It is important to remain mindful of the fact that a SR takes place within an institutional context that places certain boundaries upon its conduct. Despite the extensive use of information technology, the completion of a SR is dependent upon fallible human beings in order to manage and marshall both data and the other people involved in the review process. As such, this section provides an analysis of the inter-relationship between the managing and the review teams, and the impact that this had upon the outcome of the review.

4.1 Managing: Quantities and Teams

Whilst it may be a common refrain within any effort to review large swathes of literature (whether 'systematic' or not) that 'there is too much information and not enough time to do it justice', the manner in which the time constraints impacted upon the SR in this case study were arguably of significance. In certain respects, the tight timeframe within which the review had to be completed (so as to fit into the wider

consultative schedule of the institution) resulted in some areas of the SR process and the contents of the scope not being addressed in an ideal way. For example, utilizing two team members to review each paper in the screening process was not possible (ID3 355-362); nor was there sufficient time to review issues surrounding implementation (ID3 230-233 and 515-520). In this respect, possible solutions are reasonably well defined, and would largely be a matter of negotiating longer timeframes or commissioning another team to conduct supplementary reviews[79]. However, in other areas, it was the nature of the review process within the institution itself that was identified as having a more fundamental impact upon the manageability of conducting the review, and which was not amenable to a clearly defined solution.

The tight timeframe for the review was identified as impacting deleteriously upon important intellectual and critical processes that, whilst not amenable to being recorded within a database or represented on an evidence table, were considered to be a vital part of the process of conducting a SR. This manifested itself in the form of there being no time available in which the review team could discuss how the findings in the research could best be synthesized, nor to take a step back from the process in order to consolidate in their own minds what the evidence was pointing towards (ID3 669-680 and 1054-1064). Whilst the MT team had an appreciation of the intense time pressure that the RT were under (ID2 221-233), opinions differed as to the impact of this pressure, with certain team members viewing the pressure as productive (ID5 747-762) whilst others considered the timescale unrealistic given the breadth of the scope (ID1 312-330). Moreover, the RT felt that they had not been made adequately aware of the compressed timescale for completing the review within the wider consultative process of the institution (ID3 992-1014), nor did they feel that they were adequately consulted regarding the extension in breadth and depth of the review's scope (ID3 67-95 and 110-116).

[79] Although it should be noted that the process of negotiation itself is likely to be far from straightforward; whilst parliament's rhetorical commitment to EBP might well be high, the realities of the political environment mean that there are severe constraints upon the time and resources that can be devoted to perfecting a SR.

There existed a tension between the managing and review teams with regard to what it was realistic to do in an effort to meet the goals of comprehensiveness and adherence to a SR methodology. Again, the MT were aware of the impact that the breadth of the scope had upon the large volume of studies that either had to be screened or whose results needed to be synthesized for the review (ID1 33-38; ID2 106-112), but were themselves working within a wider institutional context. At times, this compelled them to continue pushing the RT towards comprehensiveness within a timeframe that severely limited the implementation of a rigorous SR methodology. For example, with regard to the initial search terms for the review (which resulted in over 15,000 studies to be screened for whether or not they were appropriate to be further assessed for inclusion in the review) it was noted that the MT were encouraging the RT to *add* further search terms in an effort to achieve a comprehensive search (ID3 330-342). This was a figure significantly above that which would be expected in a review with finite resources available for its completion. Moreover, this quantity precluded standard SR procedures such as searching through citation lists for papers that the search terms had missed, and double-checking of a reviewer's work by a second member of the team (ID3 364-369). On other occasions, the RT felt that they had successfully negotiated a pragmatic means of addressing a problem with a member of the MT, only to have this decision overturned by the MT at a later date (ID3 1202-1214).

4.2 Physical and emotional labour

It may appear incongruous to devote a section to physical and emotional labour in the analysis of a review that should primarily be an intellectual undertaking, and moreover one that is nominally a scientific affair detached from human emotions. However, whilst this analysis can in no way establish a causal connection between the trials of physical and emotional labour of the review teams and any particular outcome (or shortcoming) of the review process, a knowledge of this labour is arguably

important contextual material for understanding the construction of the evidence-base as a whole.

Members of both the managing and review teams used the phrase 'blood, sweat and tears' to describe the arduous experience of completing the review (ID2 758-763; ID3 1098-1105 and 1788-1802). These views were tempered with the belief that the work was important to do, and that the credibility of the institution as an organization motivated them to complete the review to the very best of their abilities (ID2 753-758; ID3 186-195). Nevertheless, team members were clear that the review process had placed severe stress upon their working roles.

The nature of the SR process set out in the methods manual provides a strongly structured approach to conducting a review that arguably contributes to the intensity of its physical and emotional labour. The review process, in particular the data extraction process, was viewed as a very important, but basically repetitive and mechanical task in which rigour was assured by adhering to the scope rather than by reviewers exercising their critical faculties (ID1 304-305; ID2 513-525). In combination with the time pressures and the sheer volume of the task, this arguably contributed significantly to the fatigue and dissatisfaction experienced by the RT at certain times during the review process (ID3 371-373 and 1059-1062). It was also noted by the RT that the institution's review process differed substantively from other systematic and non-systematic reviews which they had been involved with in terms of the time available for the team to reflect upon and discuss the findings. However, it was also opined that such review approaches could not produce the level of detail needed to establish effectiveness as would be achieved using the institution's review process (ID3 1086-1105).

Conclusion

The process of taking research apart in the SR, with a view to synthesizing the results in order to contribute to the making of policy recommendations, has been shown in this case study to be an arduous affair that is quite different from the clarity of the approach suggested by the methods protocol. This is despite the best efforts of the managing and review teams to adhere as closely as possible to the protocol; indeed, the responses of team members during interviews strongly suggests that continued adherence to a (more highly refined) protocol is viewed as the most likely way to assure the objectivity and rigour of the process. This is not to suggest that adherence to a protocol is a mechanical affair; team members acknowledged the presence of subjective elements in its utilization, but stressed that these could be adequately controlled for through implementing a rigorous management process.

The one year timeframe available for the completion of the SR needs to be placed in the context of the wider consultative process of the institution. Of this year, only three months is assigned to the completion of the review itself, the net effect being that a substantive grip upon an extensive literature had to be rigorously developed within a very short space of time. The stress of producing a review that was defensible in this timeframe was considerable, and required at times that shorter routes be found for completing the different stages of the review. The protocol required that the use of time during the review itself leant heavily towards the means of screening and appraising studies and recording the data extracted from them. This meant that a much smaller period of time was spent on the synthesis of this evidence into a form that could best be utilized by the committee in drafting policy recommendations.

Areas of ambivalence within the review process were evident too. At times this was manifested in the form of disagreements between the managing and review teams about the interpretation of the scope, and in particular to the degree to which qualitative

studies were excluded from the review through the application of the screening criteria in the methods manual. In many ways, this was an operational problem that could have been addressed in advance by checking more closely that both teams were interpreting the scope and methods manual in the same way. However, ambivalences also arose with respect to the utilization of expertise within the review itself, and in the utilization of the review by a committee. These ambivalences centred upon the concurrent utilization of research evidence (as defined in the methods manual) with the selective inclusion of expert knowledge, which team members experienced great difficulty in accounting rigorously for.

The analysis of a sub-section of the papers excluded from the review provided a mixed picture of what it might have been possible for them to contribute to the review. In certain areas, the excluded studies were clearly lacking what the review required; for example, the work was already substantively reported in studies that were included in the review, or there was a lack of evaluation of interventions in a form that could be considered rigorous. However, in other areas, the excluded studies could arguably have usefully contributed to the review; first, in providing a framework for the analysis of interventions for black and minority ethnic groups; second, by increasing understanding of the difficulties of conducting rigorous research in the field of substance misuse and of effectively implementing interventions in day-to-day service delivery; and third, by outlining different conceptualizations of the phenomenon of substance misuse in the young. Consideration of the reality of the contexts in which decisions are made by the young as to substance use, and the degree to which these could be considered 'rational' (in the sense that they are often considered in studies that posit 'intention to use' questions to respondents), could also have usefully informed inclusion/exclusion decisions regarding studies that would contribute to the review. Interrogation of all of these aspects could have been facilitated using the framework of internal, external, and construct validity, and an explicit consideration of

the balance that it was appropriate to strike between these different forms of research validity.

It is justifiable to give particular consideration here to the matter of construct validity in the design and conduct of the substance misuse review. Echoes of the moral pathological model of drug use are clearly discernible in the constructs that were utilized. For example, the focus upon the individual reflects the conceptualization of those using substances as relevant targets for professional intervention that aims to minimize or stop the use of those substances. Stated another way, the constructs utilized in the review do not include the wider determinants of health as a legitimate focus for interventions; the constructs therefore act in such a way as to exclude knowledge about these wider determinants from the evidence-base produced. The wider determinants of health are not limited to straightforward economic deprivation, as they may also act in a cultural sense through the reproduction of attitudes or behaviour that perpetuate inequalities in health. These are legitimate, indeed important, areas of enquiry, but to include them in an evidence-base requires some difficult reflection upon the adequacy of the constructs that underlie the production of the evidence-base.

Summary

The pressures involved in completing a review within a short time frame are intense, and are further compounded when endeavouring to adhere closely to a strict methodological protocol. The sheer volume and diversity of studies that were critically appraised as suitable for inclusion in the SR outweighed the ability of the methods protocol to handle them in such a way as to allow knowledge to be rigorously cumulated. It was necessary (and arguably justified) for expert knowledge to be utilized in this process, but this could not be accounted for in the terms of the methods protocol. Investigation of studies that were critically appraised as unsuitable for inclusion in the SR suggested that a number of these may have been able to contribute

substantively to a theoretical framework that could have facilitated the management and cumulation of knowledge in the review.

Chapter 6

Putting research back together again: Using a systematic review to write evidence statements and recommendations

Once the task of searching for, appraising, and extracting data from papers is complete within a systematic review (SR), the focus turns to re-assembling the evidence so produced into a meaningful whole that is capable of informing policy and practice. In short, the pieces of the puzzle have to be put back together in a new way so that the state of knowledge is moved forward. In the substance misuse review, the process utilized was one where the review teams worked together on writing evidence statements[80] which could be used by a senior committee[81] ('the committee') in order to deliberate upon and draw up policy recommendations. As in Chapter 5, the chronology of this process is outlined in order to clarify the sequence of events and the timescale in which they occurred. The manner in which evidence was marshalled and split in different ways in the teams' efforts to rigorously synthesize it is then considered, along with the actions taken by the managing team (MT) in order to present the evidence to the committee in a palatable way. Lastly, an analysis of the deliberation of the evidence by the committee considers the manner in which certain evidence was deemed acceptable, whilst other evidence was not; it is argued that an appreciation of discourses surrounding public health and substance misuse are vital for understanding the processes by which the committee deliberated the evidence.

1.0 Chronology of the process

Discussion of the first draft of the review at an internal MT meeting on July 6th focused upon a number of procedural issues (such as the distinction between study

[80] An evidence statement is a synthesis of the evidence that clearly states both its strength and applicability to the population for whom policy recommendations are to be made (National Institute for Health and Clinical Excellence, 2006).

[81] In this case study, the committee was a standing advisory body that had been convened *by* the institution (but which comprised of no members *from* the institution) to deliberate upon a range of public health issues about which the institution had conducted SRs. Membership reflected a range of health professional backgrounds, among them General Practice, epidemiology, psychology, and Health Visiting. There were also two lay members. Four 'co-optees' with expertise in the field of substance misuse contributed to the committee's discussions, but were not permitted to become involved with the deliberations *per se*.

types and the non-reporting of inter-rater reliability) and the absence of a consideration in the evidence statements as to whether study results were applicable in the context of England. It was agreed in a subsequent teleconference with the review team (RT) (July 10th) that these issues could be satisfactorily addressed. However, agreement was not reached upon other issues; first, the manner in which the review was structured (by population rather than intervention type) was questioned on the basis that substantively different interventions were grouped together. The RT responded that the way in which public services are structured is broadly around the populations identified in the review, and the review needed to be structured in this way if it was to make sense to practitioners. Second, despite the acknowledged difficulties that study heterogeneity would cause, the MT strongly advocated the use of forest plots[82] as a means of communicating effect size more succinctly than is possible with text alone. The RT responded that the studies were simply too heterogeneous to combine in a traditional forest plot, although it was conceded that *individual* studies could be represented on a forest plot in a similar way. The risk of the final review being essentially a compendium of research rather than a synthesis was acknowledged at the internal MT meeting (July 6th), but was not discussed in the subsequent teleconference with the RT (July 10th). Further points of detail were expanded upon in subsequent emails, and the finished draft of the review was submitted to the MT by the deadline of July 31st.

The RT submitted the final copy of the review to the MT on July 28th, although this was on the understanding that revisions could be made if necessary. The work of the two teams then turned to consideration of how best to present the review to the committee. At the internal MT meeting of August 8th, discussion focused upon how to present a review that contained 86 evidence statements in a meaningful way to a committee which had expertise in the field of public health, but not specifically substance misuse. The MT considered how discussion by the committee could be

[82] A forest plot shows the cumulative contribution (the effect size) of each study in a meta-analysis at a stated confidence interval (Egger and Davey Smith, 2001).

facilitated so as to increase the likelihood of reaching some firm decisions on policy recommendations. In a teleconference with the RT on August 16th, the scope of the presentation that the RT would make to the committee was agreed upon; fundamentally, the focus would be upon what the committee needed to deliberate (for example, uncertainties in the evidence, implementation issues) rather than a re-statement of the findings of the review. The presentation of further analyses in the form of forest plots was also discussed during this teleconference, the eventual outcome being that the RT did present a number of further forest plots relating to individual studies to the committee. A further teleconference between the managing and review teams was held on August 30th in order to finalize preparations for the September committee meeting; the way in which some of the initial critique of the review received from committee members would be addressed during the committee meeting was also discussed at this stage.

The first committee meeting to consider the review was held on September 4th; it was planned that the committee's deliberations would provide substantive material to allow the drafting of recommendations by the close of the meeting. However, the sheer volume of evidence for the committee to consider in conjunction with the difficulties faced in generalizing findings to the English context, meant that firm agreement upon the direction that recommendations should take was not agreed upon by the committee. At the second committee meeting (October 13th), additional presentations relating to promising interventions (in the form of revised evidence tables and forest plots) were considered, as were the draft recommendations that the MT had produced (based upon the somewhat inconclusive deliberations of the previous committee meeting and subsequent email correspondence with the committee members). This proved to be a rather frustrating meeting, the committee still not being able to reach a consensus upon the suitability of the draft recommendations.

A series of day-long 'fieldwork' events were held in Liverpool, Manchester, and Bristol during November in order to 'test' the draft recommendations with practitioners whose work brought them into contact with vulnerable and disadvantaged young people. These events were designed to elicit feedback that would allow the recommendations to be fine-tuned to better fit the realities of service delivery; they were not intended to be forums in which the review itself could be critiqued. The feedback was utilized in the third and final committee meeting (January 12th 2007) in which deliberations focused upon the precise wording of the recommendations in order that they took full account of implementation issues. The MT took responsibility for the final revision of the recommendations in view of the committee's discussion, and these were published[83] on March 28th 2007.

2.0 Synthesizing data for the review

This section presents an analysis of the way in which the managing and review teams worked together to assemble the evidence in the form of a review suitable for use by the committee. Crucially, this analysis considers how judgements were made about what evidence to synthesize when methodological criteria alone failed to provide a sufficient basis for preferring one form of evidence over another. The analysis further examines the manner in which evidence was framed in order to provide a structure for the review and how this led to the development of evidence statements of a particular form and content. The reflections of managing and review team members are utilized throughout the analysis as a means of identifying some of the tensions in the process, in particular with regard to the external validity of the review.

At the meeting to discuss the first draft of the review, it was acknowledged that the majority of the reviewers' time had been spent on extracting the data from studies rather than synthesizing this data into a form suitable for constructing evidence

[83] The guidance was published in both a detailed and 'quick reference' format. In addition, a costing summary for the interventions recommended in the guidance was produced, along with a costing template to allow the cost-effectiveness of these interventions to be calculated for individual localities within England (see: http://guidance.nice.org.uk/PHI4).

statements (RD 1791-1794 06/07/06); furthermore, the MT were wary of repeating the mistakes made in some previous reviews, where a list of research findings had been presented sequentially without being worked up into a more meaningful whole (RD 1817-1818 06/07/06; EM 10/07/06 11-14). The MT were well aware of the difficulties that the review team faced in attempting to synthesize findings even from studies evaluating the same type of intervention, for example with regard to the different criteria used to define problematic behaviour (EM 10/07/06 301-308), but were unable to offer guidance on how to resolve this issue (ID3 804-817). This resulted in the review focusing upon summaries of the results of individual studies rather than a more systematic cumulation of these results (ID3 929-935).

The MT were not neglecting the difficulties that the review team were having with synthesizing findings, but the proposed method of dealing with the problems of synthesis was not possible to do in the view of the RT. It is important to remember at this juncture that the aim of synthesizing study findings in this review was to allow them to be presented for consideration by the committee; the synthesis did *not* have to unequivocally demonstrate what guidance should be, for the purpose of the committee would be to deliberate upon the inconsistencies in the evidence-base in order to draw up considered draft guidance. As such, the aim of the research synthesis was to present to the committee the complexities of the review findings, but without involving the committee in debates over the minutiae of these findings. It was the view of the MT as a whole that the best way to present the findings to the committee was in the form of forest plots showing the direction and effect sizes of the different interventions; it was argued that if the heterogeneous nature of the studies in the review prevented a meta-analysis being performed, then showing the effect sizes of individual studies instead would be of benefit, particularly in view of their value in summarizing complex data graphically.

The idea of utilizing forest plots was first discussed at an internal MT meeting, where the problem of study heterogeneity for doing this was also discussed (RD 1814-1818 06/07/06). Nonetheless, it was proposed to the RT in a teleconference the following week that they should seriously consider utilizing forest plots in the main body of the review (RD 1913-1917 10/07/06), to which the RT replied that the studies were simply too heterogeneous for a genuine meta-analysis. By the time of the internal MT meeting the following month, it was reported that the decision had now been made not to use the forest plots in the final review (RD 2034-2035 08/08/06), but shortly afterwards the issue was raised again in a teleconference with regard to how the forest plots could be utilized in the committee meeting (RD 2143-2167 16/8/06).

2.1 Selecting and weighting evidence to synthesize

The SR process detailed in the methods manual aims to produce a synthesis of the most rigorous knowledge that can best inform policy and practice. The process is explicit in that there are clear criteria for classifying the *type* of research evidence that is found (based upon the hierarchy of evidence), appraising its *quality*[84], and assessing the *applicability* of the findings to England. It is intended that this process should clarify what research evidence should be utilized to inform policy and practice, a rigorous and applicable randomized-controlled trial (RCT) being clearly preferable to a before and after study of dubious rigour that has doubtful applicability to the English context. However, the utility of the process when research evidence is rated equally using these criteria brings forth a number of issues, especially when these studies reach different conclusions regarding the effectiveness of similar interventions. This section will focus upon one area of the review where this occurred in order to produce an analysis of how decision-making takes place where studies of equal rigour produce inconsistent results. What evidence is preferred when rating studies by type, quality, and applicability fails to clarify what evidence should be utilized?

[84] The quality of each study included in the review was graded as either '++', '+', or '-' according to whether or not the majority of the criteria in the appropriate study appraisal tool had been satisfactorily met. A '++' study fulfils most or all of the criteria, a '+' study only some (but not to the extent the study's conclusions are considered likely to change substantively). A '-' study fulfils few of the criteria, and its conclusions are considered to be non-rigorous (National Institute for Health and Clinical Excellence, 2006).

The grading of studies in SRs with regard to their type, quality, and applicability is designed to make the review process transparent. It should be clear to the critical reader of the review why certain evidence was preferred in any particular instance, but this is arguably not always the case. It would, however, take a very persistent reader an appreciable amount of time to dismantle the review sufficiently to be in a position to critique it in this way. In this section the opportunity will be taken to focus upon the weighting of evidence with regard to the 'Life Skills Training' (LST) programme and to question the transparency of the process involved.

Evidence statement 21.1 (Jones et al., 2006) reports upon the effectiveness of LST as evaluated by three RCTs rated '+' for quality and 'B' for applicability (likely to be applicable in England if appropriately adapted). Whilst short and long term reductions in substance use were not evident, there were reductions in the medium term (six months to one year). Evidence statement 21.2 reports upon further studies, the majority of which were rated substantially lower for quality, that question the effectiveness of LST when delivered in conjunction with other approaches such as parent workshops. Based upon these evidence statements, even when read in the context of the whole review, it would be quite reasonable to further discuss the LST programme at the committee meeting in order to clarify how might be used to inform policy and practice. Whilst, on the basis of the review findings, the effectiveness of LST in *combination* with other approaches is doubtful, there is strong evidence that it is effective when delivered as a *single* intervention in the *medium* term. Ideally, one would like it to be effective in the long term, but effectiveness in the medium term is still superior to that of other interventions (for example, motivational interviewing is effective only in the short term (evidence statements 52.1, 52.2, and 53.1)).

In the committee meetings, the focus of the committee was steered away from the evidence statements that reported the effectiveness of the LST programme. There was also concern within the managing and review teams that the economics team had

modelled cost-effectiveness on the basis of the LST approach and that the committee would, mistakenly, want to recommend LST on this basis (ID3 1464-1475). Why did this significant degree of opposition to the LST approach exist when, working from the evidence statements, it might be reasonably expected that LST could contribute to the making of recommendations? Arguably the opposition was rooted in a report that the MT had suggested that the RT should take into account (RD 363-370 27/04/06). This report, conducted for the Scottish Executive Drug Misuse Research Programme (Coggans et al., 2002), provided a substantial critique of the LST programme. The report argued that the extensive body of research on LST showed that its effectiveness for stopping or decreasing the use of illicit drugs was negligible (and certainly far less than is portrayed in the promotional literature for the intervention), that the costs of delivering the programme were high, and that fidelity of delivery (often problematic when working with high risk populations) was crucial if the limited effectiveness of the intervention was to be delivered. The rigour of some of the evaluations conducted or co-authored by the designers of the LST programme were also called into question.

Coggans et al.'s (2002) critique was arguably important for the conduct of the review. However, it is important to note a number of issues relating to its use. First, the report was not found using the database search conducted for the review; it required an expert professional network (existing outside of the documented SR process) in order to highlight the report, argue the case for its importance, and bring it into the synthesis of knowledge being conducted by the review team. Moreover, the role that the report played in shaping the management of the committee is entirely unreferenced in the review. The reader of the review would be unaware that there was contention over the effectiveness of LST, and would also be unaware of why the approach was not favoured in the deliberations of the committee.

Second, because of this utilization of both a SR approach *and* the influence of the expert knowledge of a professional network surrounding the review and the

institution conducting it, the review finds itself struggling to justify contradictory statements. Thus, whilst evidence statement 21.1 cites the three strongly-rated studies offering guarded support for the effectiveness of LST (Griffin et al., 2003; Smith et al., 2004; Vicary et al., 2004), the use of the review in the deliberations of the committee needed to be managed in such a way as to not make this support too clear. Critique of the review is inhibited by the lack of acknowledgement of the role that Coggans et al.'s (2002) critique played in shaping the deliberations of the committee; moreover, what is the critical reader of the review to make of the studies cited in evidence statement 21.1? Should they not be weighted as highly as is suggested by their strong quality and applicability ratings? In other words, is there other substantive evidence that should be included in the statement? If so, what is the nature of that evidence? If it is Coggans et al.'s (2002) non-systematic review, upon what basis is *this* evidence included in the review, but not other non-systematic reviews on other subject areas? Do the studies conducted after the publication of Coggans et al.'s report (Griffin et al., 2003; Smith et al., 2004; Vicary et al., 2004) address the issues identified in that report and thus, effectively, supersede that report's findings? Or do they fail to address these issues, in which case, is an assessment of study quality based solely upon methodological criteria adequate for the purposes of producing a rigorous synthesis of knowledge?

None of the above should be taken as implying that the utilization of Coggans et al.'s (2002) report was unjustified. The issue is with regard to the SR method utilized. If it is necessary to utilize expert professional networks in order to inform the synthesis and deliberation of knowledge in a SR, the rigour in a SR that is claimed to be attained through the transparent documentation of database searches and decision-making regarding the inclusion of papers is substantively brought into question. To be clear, this is not to suggest that the substance misuse review was in some way guilty of malpractice; rather, it is argued to be a strong example of the *necessity* in SRs of utilizing networks of expert professional knowledge, despite the outward appearance of

reviews being one where this expert knowledge is portrayed as of doubtful objectivity and hence to be avoided.

2.2 Options for synthesizing evidence from included studies in a different manner

This section focuses upon just one evidence statement in the review in an effort to explore how the studies included in the review might have been utilized differently and so as to better inform the deliberations of the committee. Evidence statement thirty-two[85] reports the effectiveness of family-based interventions for improving secondary outcomes (considered to be protective factors against substance misuse) in the short-term in African-American families. The evidence statement drew primarily upon the results of three RCTs rated as '+' for quality (reported in four papers: Hogue et al., 2002; Spoth et al., 2003; Brody et al., 2004; Brody et al., 2005) and was further endorsed by the results of two controlled before and after studies (both rated '-') (Bruce and Emshoff, 1992; Aktan, 1999) and one before and after study (rated '-') (Emshoff et al., 1996). The conundrum faced by the committee with regard to this evidence statement was that the evidence presented to them was strong, yet the review had also rated the 'applicability' (to the English context) of the studies as 'C' (applicable only to populations or settings included in the studies). In effect, this rating states that the findings are not applicable in the context of England. Whilst the review contained a brief discussion of the difficulties of generalizing from the results of United States of America (US) studies to England, the committee found that this did not facilitate them to make judicious inferences from the evidence in this example in order to make recommendations.

The analysis presented in this section therefore returns to two of the highly-rated studies (Hogue et al., 2002; Brody et al., 2005) and the study rated as the

[85] This evidence statement was purposively selected for further analysis for two reasons. First, the setting of the studies cited (all were conducted in the US) highlights a recurring problem throughout the utilization of SRs for policy making; the issue of external validity, or how to apply (in this instance) high-quality evidence from the US to England. Second, the evidence statement focuses upon effective interventions in black and minority ethnic groups, an area about which there is acknowledged to be a lack of knowledge.

weakest (Emshoff et al., 1996) in order to explore whether or not these papers[86] contained further details that would have usefully informed the committee in drawing up recommendations for policy and practice in England. In doing so, the question being asked of the SR process is with regard to internal validity; does placing an emphasis upon grading studies according to their internal validity and extracting data solely with regard to outcome data result in the neglect of much-needed information about external validity?

Brody et al.'s (2005) paper acknowledges that their evaluation of the 'Strong African American Families' programme contained no explicit consideration of the generalizability of the study- results to African American families residing in areas outside of the study focus, or who had a markedly different socio-economic status. In this respect, the potential utility of the study for informing policy making in England does indeed appear slight. However, the paper continues in discussing the profile of the 'working poor' African American families that were the subject of the research; in doing so, it clearly sets out wider environmental factors that are posited to drive the propensity to misuse substances by the young. These factors include limited employment opportunities, a lack of recreational facilities for youths, and limited provision of physical and mental health care. In this sense, these families have substantive similarities with those in England who are socially excluded; it would arguably be wholly defensible to make a considered generalization of the results to the English context.

Hogue et al.'s (2002) paper provides a very detailed evaluation of the 'Multidimensional Family Prevention' programme. Again, the authors are tentative

[86] The three papers were selected purposively on the basis of the following criteria. First, to enable exploration of how highly-rated studies *and* low-rated studies might have been able to contribute other evidence to the review; it was considered that two studies rated '+' and one rated '-' would provide a reasonable representation of this range. There was an even split in the '+' studies in that only two programmes were evaluated between these four studies; a coin was tossed to determine which of each pair of studies would be analyzed further. The '-' study was selected on the basis of convenience; an electronic version was available, which was not the case for the other two studies rated '-'.

about the external validity of the study, but again the utilization in the study of a risk profile (consisting of factors such as chronic school truancy, history of delinquency, and previous drug use) suggests that the authors consider that there are sufficient similarities in the roots of such behaviour that would allow them to be addressed using interventions based upon the same principles. Perhaps more importantly for the purposes of the committee, the study found that the positive impact of the programme upon secondary outcomes such as self-concept and school bonding were robust across all of the research subjects, regardless of their sex, age, or their behavioural problems upon recruitment to the study. Whilst this does not mean that the study findings are automatically applicable to the English context, it does provide evidence to suggest that the findings could be utilized rationally.

Emshoff et al.'s (1996) paper is more problematic with regard to external validity. The paper itself contains less detail than either Brody et al. (2005) or Hogue et al. (2002), making it less clear exactly what the intervention involved and in what way the findings might be generalizable. The authors do note that the exact content of the intervention was documented, and this would presumably be available on request from the authors, but doing so in the context of a time-pressured review is likely to be very difficult. The authors again note the context in which the intervention took place, this being an inner-city neighbourhood characterized by poorly resourced public services and unemployment, and the manner in which this environment drove the disaffection that placed youths at high-risk of substance misuse. As previously, the suggestion can be made that this is a picture of social exclusion that can also exist in England; however, the research in the paper does not provide evidence that clarifies the mechanisms at play and which would facilitate generalization to the English context.

Although the lack of precise details regarding the intervention studied in Emshoff et al. (1996) poses problems for using the research to make recommendations for English policy, there can equally be problems in the conduct of the review when

more extensive details are provided. Arguably, the problem here relates to the need in the review to synthesize knowledge; doing so requires that similar interventions (in this instance, 'family-based' interventions) be grouped together in an effort to make evidence statements that provide a rigorous overview of research findings. The danger in doing so, which is particularly acute in a review of this scale completed in a short time period, is that the extraction of data from the original papers is insufficient for the subsequent synthesis of knowledge.

The judgement made about the optimum level of data to extract is a perilous one, and the danger exists that if the data extracted is insufficiently detailed, or provides the 'wrong' sort of data, then substantively different interventions can end up being treated in the review as fundamentally the same. For example, in the substance misuse SR, evidence statement 32 justifiably classifies both the 'SUPERSTARS' (Emshoff et al., 1996) and 'Multidimensional Family Prevention' (Hogue et al., 2002) programmes as 'family-based' interventions. However, the resources required to attain fidelity in the delivery of these interventions differ radically. The 'SUPERSTARS' programme consists of artistic events delivered to groups over the period of about ten weeks; whilst precise time commitments or costings are not provided, the impression given is that the programme could be effectively delivered using existing community organizations and some protected time for a community worker in order to organize and oversee the artistic events. In comparison, the 'Multidimensional Family Prevention' programme requires counsellors trained to a Masters or Doctoral level, who furthermore undergo training over a period of four months in order to learn how to deliver the interventions (fifty hours of seminars, plus thirty hours of supervision reviewing recorded counselling sessions and individual supervision of two counselling 'pilot case'); during the delivery of the intervention (which could extend over many months, depending on the progress of the families concerned), the counsellors received three hours per week of individual supervision from one of the study authors. This is arguably vital information for making inferences from the knowledge to the

English context, as without it cost-effectiveness has to be calculated on the basis of all 'family-based' interventions costing the same to deliver. In this instance, this is manifestly not the case, and important knowledge that would have informed the committee's decision making process was lost.

Finally, returning to the issue regarding external validity raised earlier regarding the wider environmental drivers of substance misuse in the young, a key question can be asked of the included papers which resonates throughout the entire substance misuse review. All of the papers analyzed in this section (Emshoff et al., 1996; Hogue et al., 2002; Brody et al., 2005) focus upon 'high-risk youth' in locations where the *socio-economic environment* is acknowledged to disadvantage them in multiple ways, yet all of the interventions focus upon the *individual and their family*. Whilst the interventions are evaluated as being effective in terms of secondary outcomes in the short-term, policy making is about both the short- and long-term. Whilst these papers do not establish that the socio-economic environment *causes* substance misuse, nor quantify the extent of its contribution to substance misuse in the young, there is a strong recognition that disadvantageous conditions prejudice certain youth's life chances. This begs the question of why the studies are not focused at the level of the environment and the manner in which it impacts upon substance misuse in the young. Moreover, in terms of the SR, it might justifiably be asked why the review was structured in such a way as to focus upon interventions at the level of the individual or family rather than upstream at the level of the environment.

2.3 Structuring the evidence-base

The work of re-presenting the evidence-base in document form for presentation to the committee involved the submission of drafts of the review being submitted by the RT in order that the MT could give feedback. This feedback first took the form of mundane but important proofreading; for example, sifting through the review to pick up on repetition between the text and what was presented in the tables, and challenging

inconsistencies or ambivalence in the text (EM 10/07/06 16-18). Second, feedback involved more fundamental structural guidance that was designed to make the document usable by its intended audience, that is, stakeholders and the committee. This structural guidance could be similar to that which would be given by any critical reader who has the advantage of looking afresh at a piece of work; thus, suggestions were made by the MT regarding the presentation of data in tabular form (for example, listing studies alphabetically by author rather than by study type (EM 10/07/06 138-144), combining certain tables rather than presenting them separately (EM 10/07/06 28-30)) and in the consistent use of terms throughout the review, for example where authors had utilized terminology in different ways in their respective review sections (EM 10/07/06 39-41 and 46-48). In this manner the MT endeavoured to provide constructive feedback that would facilitate the RT to revise the document so as to make it internally consistent, usable by both a committee and a wider audience, and possible to defend against critique by those parties.

The style of feedback outlined above has many similarities with that traditionally given by an editorial team or through peer review. However, the nature of the review being part of a wider consultative and deliberative process with tightly defined deadlines meant that there were substantive areas where the feedback from the MT differed from the more conventional editorial or peer review approach. The MT strongly suggested to the RT that their having structured the review by population (i.e. general 'at risk', black and minority ethnic group, and so on) meant that, despite the further sub-headings for different intervention types, wide ranges of interventions were subsumed under each of the headings. Furthermore, this structure meant that comprehending the review was not as straightforward as it could be (RD 1770-1776 06/07/06). The RT argued that the mode of delivery of drug education services meant that practitioners tended to work predominantly with the population groups as outlined in the review structure, and it was thus more logical to present the evidence in this way in order to make it more usable. However, the RT agreed to try and revise the structure to address

the concerns of the MT, but warned that the diverse nature of the interventions assessed would make splitting the review into discrete intervention types highly problematic (RD 1909-1911 10/07/06; see also ID3 804-817). When reflecting upon the review process, the RT identified the large number of papers included in the review and the strictly limited time available to investigate particular approaches in more detail (in order to better understand them and propose a more developed 'intervention type' structure) as key drivers of the review being kept on a broader level relating to intervention types (ID3 827-847).

The RT's efforts to restructure the review resulted in a limited breaking down of the outcomes of certain intervention types into more finely distinguished groupings, rather than anything more fundamental. In effect, despite the efforts of the MT to persuade the RT to restructure the review along population lines, the RT simply exercised their veto as the team actually writing the review and who, in their professional judgement, could see no strong reason for fundamentally restructuring the review in the very limited time available. Arguably, there was a tension between the demands of the MT and what it was realistic to do with the large and diverse volume of evidence using the methods proposed and in the timescale available. This manifested itself in the pressure from the MT for the review to more precisely identify what interventions were effective for whom (and in what contexts) (EM 10/07/06 151-155, 170-172 and 202-204; EM 19/07/06 92-105). This clashed somewhat acutely with the RT's efforts to make judicious groupings of intervention types in a bid to conduct at least a limited synthesis of the diverse evidence which they had to draw upon.

2.3 Writing evidence statements

This section will examine the requirements made of the RT made by the MT with regard to the content and format of the evidence statements in the review. Consideration will also be given to the different perspectives which the two teams brought to the process and the manner in which this impacted upon negotiations

regarding the content and format of the evidence statements. The manner in which studies were framed within evidence statements will also be considered with regard to the different interpretations that were made of research findings.

A key criticism of the evidence statements in both the first and second drafts of the review submitted by the RT to the MT was that they did not answer questions around implementation issues that had been highlighted in section 2.1 of the review's scope. These issues covered areas such as geographical location (for example, would what works in Holland work in the England?), the nature of the practitioner's profession (for example, is an intervention equally effective delivered by a school teacher as by a drugs worker?), and the acceptability of the intervention to different target audiences (for example, do boys and girls react differently to the intervention?). It is acknowledged in the methods manual that, in public health, a systematic procedure for assessing the robustness of a study's findings in different contexts (i.e. its external validity) has yet to be developed; understandably, this left the RT struggling in many instances as to what basis they could legitimately claim that evidence was applicable or not to particular populations in the English context. In short, there was no explicit theoretical framework stating that (for example), certain contexts or populations were considered to share sufficient key characteristics that would allow the findings in one to be generalized to the other. For some members of the MT, the task of assessing external validity was the responsibility of the committee (ID1 381-409), whilst for others the situation was more cleanly delineated, with the four 'levels of applicability' effectively being just two; applicable or not applicable (ID2 689-705). It is interesting to note that this approach contrasts markedly with the difficulties which the RT felt they faced in attempting to assess external validity, and in their awareness that even interventions in English settings may have been delivered in quite different political and social environments to those pertaining today (ID3 1339-1342).

The RT also experienced difficulty with regard to the equivocal nature of many of the findings. Although synthesis in the form of a meta-analysis was not considered possible because of the heterogeneity of the data, the RT considered the interventions in certain studies to have sufficient similarities to allow a narrative synthesis of the findings. However, the lack of consistency in the findings of different studies often led to the situation where evidence statements concluded that there was 'insufficient evidence' to make a judgement one way or the other with regard to intervention effectiveness. For the MT, this could be addressed in two ways; first, by a more precise delineation between different types of interventions, thereby allowing more precise evidence statements to be made (albeit at the cost of producing further evidence statements in a review that already contained a very large number) (EM 19/07/06 40-54), and second by selectively exploring the equivocal areas in an effort to elucidate the factors that account for the differences in effectiveness (EM 19/07/06 56-62). Whilst the RT did not disagree in principle with these proposals, the time pressures of completing the review precluded them being followed up in a substantive manner. Again, a difference in perspective between the managing and review teams was apparent, with the MT viewing the process of writing evidence statements as primarily a technical process utilizing the data extraction tables (ID1 447-455) and the RT team reflecting upon the intense difficulties they experienced in trying to synthesize a very diverse body of evidence into the format required in an evidence statement (ID3 788-817).

The precise wording of the evidence statements was also a contentious issue between the two teams. In order to present the evidence statements in such a way as to facilitate discussion and the drafting of recommendations at the committee meeting, the MT strongly advocated that evidence statements should be worded to reflect what it was postulated (on the basis of the evidence) *would* happen if that intervention were implemented, rather than simply reporting what *did* happen when the intervention was implemented in the context of the original research (EM 04/08/06 83-89). The actual

change in wording in the evidence statement is minimal, as the following example shows (see EM 04/08/06 65-77):

Original RT evidence statement:

There is evidence from 4 RCT+ that school-based life skills training/ resistance skills interventions reduced tobacco and alcohol use compared to no intervention in populations of mixed ethnicity in the short, medium and long term.

MT's revision of evidence statement:

There is evidence from 4 RCT+ *to suggest* that school-based life skills training/ resistance skills interventions *can* reduce tobacco and alcohol use compared to no intervention in populations of mixed ethnicity in the short, medium and long term.

However, the RT were quite reluctant to produce all of the evidence statements in this manner, as to do so involved interpretations as reviewers that they did not feel qualified to make; in short, they felt that they could present a rigorous review of research evidence, but that this synthesis (in which the strength of evidence, it's applicability, and the precise findings were presented) could only be presented in this form, i.e. a synthesis of what the findings *were*, not what effect implementing them would likely *have* (EM 04/08/06 130-163 and 32-53; ID3 901-914).

The outcome of the discussion between the managing and review teams regarding the form of words used in the evidence statements was that the RT conformed with the requirements of the MT, despite continuing to express their reservations as to the appropriateness of the wording. However, framing the analysis of the outcome in this manner risks giving the impression that the MT were encouraging the RT to claim far more in the evidence statements than was warranted, when arguably the situation is the other way around. The MT's wording, whilst it involves extrapolations from the results of the studies themselves, is careful to note that the

result is not clear cut and is unlikely to apply in all places at all times. In contrast, the RT's wording, whilst technically only reporting the results of studies, in effect claims far more; it is quite definite about the effects of the interventions and moreover identifies no limitations over space and time of these effects.

3.0 The organization of the committee by the managing team

The difficulty faced by the MT with regard to the committee was primarily with regard to how to manage a group of professionals in order to facilitate constructive discussion, but also how to focus this discussion in such a way as to lead towards rigorous and defensible policy recommendations being made. The actions taken by the MT indicated that it was not considered sufficient to simply present the review to the committee and ask them to deliberate upon its findings. Instead, a far more pro-active approach was taken. This involved the precise planning of the meetings' agenda, the investigation and resolution of contentious issues prior to the meetings, and the preparation of further analyses of data in order to facilitate debate by the committee.

The overarching aim of the planning of the agenda for the committee meetings was to avoid the worst-case scenario of committee members debating various points at length without making any significant progress towards drafting recommendations. The MT, in consultation with the RT and the Chair of the committee, therefore worked towards presenting the review in such a manner as to best focus the thoughts of the committee on the task at hand (the drafting of recommendations) rather than allowing discussion to range across various other issues regarding the review, most pertinently the review methods (RD 08/08/06 2092-2093). The content of the introductory presentation by the RT at the meeting was intended to perform the role of setting the agenda for the committee's discussions; the content was negotiated in an uncontroversial fashion between the two teams, to include not solely the interventions that had clearer evidence of effectiveness, but also areas where there was greater uncertainty that required the active deliberation of the committee. These were areas

where the evidence of effectiveness was equivocal, there were uncertainties over the applicability of the evidence to the English context (or its fit with current policy and practice) and the potential for unintended (and adverse) effects resulting from the intervention (RD 2062-2066 and 2095-2100 08/08/06; EM 15/08/06 12-34; RD 2281-2296 30/08/06).

Steering the topic for discussion in the committee meetings away from the review methods and towards the drafting of recommendations was also attempted by addressing contentious issues raised by the committee members before the meetings took place. Committee members discussed the review prior to the meetings using a secure web-based discussion board hosted by the institution, and this not only allowed the committee members to critique the review or seek clarification on particular issues, but also enabled the MT to investigate and address particular issues. Such work was viewed as key in view of the importance of tightly focusing the discussion in meetings on drafting recommendations rather than debating the evidence-base (the SR) that was intended to inform the making of recommendations. As a researcher, it was not possible to gain access to this discussion board, thereby precluding a thorough analysis of its proceedings. However, it is possible to utilize the discussion surrounding some of the issues in meetings between the managing and review teams to increase understanding of the process by which these comments were addressed.

In certain instances, the nature of the critique of the review was one which the managing and review teams had been struggling with throughout the review process; the evidence statements regarding the studies reviewed could be framed in a number of ways in an effort to adequately differentiate populations and intervention types. However, the wide scope of the review and the efforts made to clearly delineate between the effectiveness of different types of interventions had meant that the number of evidence statements made increased to eighty-six, a quantity that some members of the committee felt made the review unusable. Whilst the number of evidence

statements could be reduced through introducing a greater degree of generality into their wording, this would be at the expense of identifying with precision the exact type of interventions that were effective (for example, see RD 2582-2603 01/09/06). No solution was identified for this conundrum.

In other instances, critique of the review was challenged and effectively rebuffed prior to the meeting in order that the issue was not raised again at the meeting and the deliberation of issues relating to the drafting of recommendations was not postponed. For example, the absence of effect sizes and the lack of differentiation between outcome measures (self-report or standardized) utilized in the review was raised. This critique focused on the manner in which this could confound the committee's deliberations – how could they know whether they were deliberating upon a study that showed a large effect size utilizing a standardized measure (i.e. of substantive importance) and one that showed a small effect size utilizing self-report (i.e. potentially meaningless)? (RD 2224-2236 30/08/06). The RT provided a written defence of the review in this respect, and a member of the MT personally discussed the issue with the committee member who had raised the issue (RD 2327-2341 30/08/06). However, it should be noted that a consensus was *not* reached within the MT regarding this issue. One MT member contended that the critique was justified and that it was the lack of integration between the intervention effectiveness review and the cost-effectiveness review that had exposed this weakness (RD 2343-2357 30/08/06).

The MT also made further efforts to present the findings of the review in a more usable form to the committee by directing the RT to produce 'forest plots' that showed the direction and effect size of various interventions. These were not forest plots in the sense usually understood in a meta-analysis; rather than presenting study results cumulatively in order to produce a composite effect size with a tighter confidence interval, the plots simply presented results consecutively. However, it was considered necessary to produce these forest plots in an effort to facilitate the committee's task of

distinguishing between the often equivocal results presented, and to bring into better focus what the real-world impact of such interventions could be (EM 15/08/06 36-47). Discussion between the managing and review teams regarding the use of these forest plots revolved around the two issues of the difficulty of splitting up data that was presented in composite form in the original studies and the risk of presenting an unwarranted degree of certainty in the forest plots. Whilst neither of these issues was sufficiently clear-cut to allow definitive agreement upon a solution, a consensus was reached on the basis that the forest plots would be limited by the nature of the presentation of the original data in certain studies and that it was permissible to annotate the forest plots in order to highlight their limitations (RD 2143-2162 16/08/06). The manner in which the committee would use the forest plots was also highlighted; it was contended that committee members would interpret the forest plots rather than simply accepting what they presented uncritically (RD 2164-2168 16/08/06).

4.0 The use of evidence by the committee: What was considered to be valid evidence?

The starting point for exploring what the committee considered to be valid evidence is one where the large volume of evidence with which they were presented had already been subject to a significant degree of sifting. This sifting had taken place in an effort to produce a review of quality that presented the best evidence available (see Chapter 5). The analysis here will thus focus upon how discussion within the committee took place about the evidence which was presented and how the arguments that were advanced regarding the state of the evidence-base were responded to.

In an effort to work towards making recommendations that could be tracked back to the evidence, the committee endeavoured to identify discrete interventions for which there was strong evidence. In doing so, an analogy with medical interventions was drawn – if the dose of a drug is effective for a limited time period only, *ceteris paribus* that dose is simply repeated in order that a desirable outcome is again

attained. It was advocated that such an approach should also apply with regard to the interventions being considered to reduce substance misuse (RD 2705-2721 04/09/06). Disadvantaged and vulnerable young people are proposed to, in effect, be 'inoculated' against the harms resulting from substance misuse through receiving the appropriate intervention. It would be inaccurate to state that all of the committee's deliberations were underpinned by this construct, but it did arguably serve to frame a substantive portion of the discussion. Moreover, the motivation of the committee to firmly identify these effective, discrete interventions drove arguments for focusing upon evidence regarding short-term effectiveness; the point was made that the difficulties involved in rigorously following-up a high-risk population at periods of greater than around three months (RD 2792-2795 04/09/06) justified a focus upon short term interventions.

The committee had difficulty in balancing their desire for making recommendations based upon discrete interventions with maintaining what they themselves viewed as a rigorous 'evidence-based' approach. Whilst it was considered a useful proposition to investigate certain interventions further, even when these had been evaluated using less rigorous research designs, the committee could not reach a consensus on how to do this in a transparent manner. The concern was that whilst (for example) listing the studies in order of their effect sizes would highlight certain approaches, this would not allow the considered investigation of promising approaches that had been evaluated using less powerful study designs; however, if the committee deliberated upon and selected certain interventions to investigate further, upon what basis were they doing this? Would the committee be accused of bias in their selection? (RD 2797-2821 and 3076-3083 04/09/06). The outcome was that the MT selected certain interventions that they judged to warrant further investigation, and about which the RT produced further, more detailed forest plots for the subsequent committee meeting.

The drive within the committee was not limited solely to identifying effective, discrete interventions; notably, there was pressure from certain members for the evidence to be investigated further in such a way as to elucidate the causal mechanisms by which the interventions were effective[87]. It was proposed that if the committee could attain an understanding of these mechanisms, it would substantively facilitate both their ability to make accurate recommendations that had resonance with practitioners *and* to focus the committee members' minds upon some of the complexities of what it was they were attempting to achieve with respect to substance misuse (RD 2984-2987 04/09/06; RD 3507-3522, 3553-3556 and 3644-3647 13/10/06). This proposal is intriguing, for it indicates that at least some of the committee members were strongly aware that the constructs being utilized by the committee were insufficiently elucidated. Nevertheless, a consensus was not attained on this argument despite it being raised on a number of separate occasions by different committee members. This was not because of any strong objection to it, but rather because the committee seemed unable to find a way in which to pursue the objectives of the proposal[88], particularly when the issue of the difficulty of extracting such information from the original studies was raised by the RT (RD 3521-3522 13/10/06).

The issue of making inferences from the evidence in order to make recommendations for English policy also posed significant problems for the committee. In short, the committee had difficulty in establishing a basis upon which they could rigorously justify taking the results regarding effectiveness achieved in one context and stating that they would be equally effective in another (RD 3008-3010 and 3088-3089 04/09/06). Input from the co-optees suggested that the issue was complex; in certain cases, for instance with regard to drug treatment and testing orders in the US, the

[87] The similarity here between realists' call for the investigation and testing of mechanisms (rather than a reliance upon regularities in association) and the desire of some committee members to make recommendations based upon an understanding of causal mechanisms, is notable.

[88] It is arguably justifiable here to infer that the difficulties encountered in trying to take the issue forward were due to the committee working within the dominant discourse of EBP, where the focus is upon outcomes rather than the understanding of mechanisms. However, as noted in Chapter 4 (section 3.1), access to recordings of the meetings was not obtained, thereby precluding explicit discourse analysis that could support or refute this argument.

different socio-political environment posed serious obstacles to generalizing evidence of effectiveness to England (RD 3113-3118 04/09/06). However, in other cases, the experiences of certain disadvantaged youth in the US could have a striking resonance with the experience of those in England, meaning that there were strong reasons to believe that the intervention would be transferable (RD 3225-3231 04/09/06). The point being made here is that if these details and considerations are not pulled out from the original studies for the committee to deliberate upon, they have little upon which to base their decision-making with regard to the external validity of study results.

The committee experienced further difficulties in getting at the information that they wanted with respect to its presentation. The substantial volume of information in the review was generally considered to make the committee's task more difficult, and a number of suggestions were made for how the information could be presented differently so as to facilitate the committee's deliberations (RD 3090-3092, 3141-3148, 3199-3206 and 3212-3217 04/09/06). Arguably these suggestions were prompted by the unease which certain committee members felt regarding the content of the committee's deliberations, namely that it was largely taking place without specific reference to the evidence statements contained in the review (RD 3011-3012 and 3057-3083 04/09/06) and as such was falling some way short of being an 'evidence-based' discussion. In addition to the difficulties experienced with the volume and presentation of the evidence, some committee members argued that they required a far more developed introduction to the field of substance misuse and the policymaking priorities within it, if they were to be able to deliberate upon the evidence knowledgably (RD 2733-2742 04/09/06; RD 3553-3556 13/10/06). Although not stated in these terms in the course of the committee meetings, arguably this again shows how some committee members were aware that the framing of the committee's deliberations was simply assumed rather than robustly constructed and defended.

In view of the difficulties experienced by the committee in comprehending such a large volume of evidence (RD 3035-3036 and 3062-3064 04/09/06), a number of its members suggested that the committee was making insufficient use of the co-optees' expertise (RD 2993-2995 and 3141-3146 04/09/06). It was unclear from the arguments advanced whether it was considered that the committee was at fault for not utilizing the co-optees' expertise, or whether the co-optees were at fault for failing to address their contributions to the concerns of the committee. Whether it was primarily the committee's or co-optees' 'fault', or whether it was some combination thereof, it is instructive to consider the contributions which co-optees did make to the discussion and which arguably were either overlooked or not considered sufficiently important for the committee to deliberate upon.

Co-optees highlighted a range of issues pertaining to the construction of the evidence-base: the uncertainties of knowledge in the field given the difficulties of evaluating interventions over extended periods of time (in particular the limited utility of RCTs); the risk of basing recommendations upon the outcomes of well-funded programmes which have a strong tendency to be the ones which are evaluated (this being insufficient reason to look past other 'weaker' evaluations that may be of greater relevance); the preponderance of school-based evaluations failing to measure the impact upon disadvantaged and vulnerable youth who may not be present at school; and the potentially decisive impact on effectiveness of the relationship between the person delivering and receiving the intervention, rather than the precise content of the intervention (RD 2649-2652, 2659-2661 and 2723-2725 04/09/06; RD 3526-3527 13/10/06). In many ways, these issues were fundamental to the construction of the evidence-base, and if the concerns of the co-optees regarding them is justified (as it might reasonably be argued they are) they serve as a strong critique of a review method which utilized a hierarchy of evidence that prioritized internal validity; failed to consider, or develop an argument for or against using, the dominant form (and assumptions) of research funded within a particular paradigm; did not develop a

strategy for cumulating evidence relating to groups of concern that it is acknowledged are hard-to-reach; and did not adequately consider the confounding nature of (for example) educators' individual personalities or professional skills upon the results obtained in evaluations. On a number of these points, the RT responded that the level of detail contained in the studies was insufficient to address the concerns raised, but the ramifications of this for making recommendations were discussed no further by the committee, nor was the role played by the review method in constructing the evidence-base in this way discussed.

Co-optees also raised a number of substantive issues with respect to the wider determinants of substance misuse. Broadly, these could be grouped under the term of 'social exclusion', and included the impact of factors such as unemployment, poor access to public services, stigma associated with substance use, differences between urban and rural locations, housing policy, poverty, and the inter-generational transmission of substance use (RD 2676-2681, 2744-2749 and 2751-2755 04/09/06). The Advisory Council on the Misuse of Drugs document, *Substance Misuse and the Environment*, was identified as a key summary of these issues and starting point for deliberations with regard to these wider determinants (RD 2744-2749 and 3041-3046 04/09/06). One co-optee succinctly argued that facilitating change in these wider determinants could far outweigh the degree of impact that educational interventions could have, the implicit suggestion being that the committee were fundamentally misguided if they thought that deliberating the effectiveness of educational interventions would genuinely address the causal mechanisms leading to substance misuse (RD 2676-2678 04/09/06). Again, these contributions to the discussion by the co-optees were arguably of substantive importance, but given that they did not resonate with the evidence as it was presented in the review to the committee, the issues were passed by and not deliberated upon.

Conclusion

The final policy recommendations of the committee, which it is expected that practitioners in England will utilize in their delivery of services, are presented (in addition to background information) in full in a document available online (National Institute for Health and Clinical Excellence, 2007) and in 'quick reference' format in hard copy. The five recommendations are summarized here in order to communicate the essence (rather than the detail) of what they contain. Summarizing in this way is proposed to better facilitate an understanding of the recommendations within the context of an analysis of the process of constructing the SR which informed their production:

Recommendation 1: Strategies to address substance misuse in the young should be developed that are cognizant of local population profiles, and which clearly define how these services will be delivered.

Recommendation 2: Practitioners whose work brings them into contact with young people should use screening tools to assess those who are at risk of misusing substances; those at-risk should be supported or referred (as appropriate) to other support services.

Recommendation 3: Practitioners should provide structured family-based programmes of support to those at-risk; these should include motivational interviews (at least three per year), assessment of family interaction, and parental skills training.

Recommendation 4: A particular sub-group of at-risk youth (aged 10-12, who consistently behave in a disruptive manner) should be offered group-based behavioural therapy delivered by specially trained practitioners.

Recommendation 5: Practitioners trained in motivational interviewing should offer this intervention to youths who are problematically misusing substances. The interviews should promote reflection upon a range of health, social and legal issues and set goals to stop or reduce their misuse of substances.

Viewed as a whole, these recommendations are clearly focused upon interventions that are delivered at the level of the individual and their family rather than at the level of the environment, the upstream source of the phenomena that it may be argued contribute substantively to problematic substance misuse. This is not to suggest that the committee and institution deliberately discarded evidence relating to upstream interventions, nor that they remained wantonly ignorant of the important role played by the environment. However, this downstream focus *does* strongly suggest that the SR method utilized (which prioritized the internal validity of studies, the criteria for which was most likely to be met by studies involving tightly-defined interventions with individuals) produced a body of evidence that largely by-passed what may have been useful knowledge about the effectiveness of upstream interventions. Whilst it might indeed be an important contribution to policy and practice to identify the effectiveness of a particular intervention for a high-risk sub-group (as in Recommendation 4), one might argue that this should be placed in the context of a wider appreciation of the impact of the environment on the phenomenon of substance misuse.

The difficulties that the committee experienced in attaining a grasp upon the evidence in the SR is arguably reflected in Recommendations 1 and 2. In contrast to the other recommendations, where the evidence statements on which they are based are listed, these first two recommendations are stated to be 'inferences derived from the evidence'[89]. If the logical inconsistency of this claim is bypassed (an inference can

[89] An 'inference derived from the evidence' is made where "... a recommendation is not taken directly from the evidence statements, but is inferred from the evidence" (National Institute for Health and Clinical Excellence, 2007, p.24).

only ever be made on the basis of evidence and through the use of reasoning based on that evidence), it may be seen how the SR struggled to construct an evidence-base that could inform the setting of policy recommendations that would shape the day-to-day delivery of services. Whilst there is arguably little that can be criticized in these first two recommendations, it is difficult to see how quite how they were derived from the SR itself as opposed to simply being the informed articulation of policy by an experienced committee.

Recommendation 5, as already noted, is an example of an intervention at the individual rather than environmental level, and as such does not address the wider determinants of health. However, what is particularly interesting about this recommendation is that it is based upon evidence of effectiveness in the short term (1 to 6 months post-intervention) rather than what amounts to quite equivocal evidence about effectiveness in both the medium and long term. Arguably, short term effectiveness may indeed be worthwhile pursuing, but the question remains as to why motivational interviewing was preferred over any number of other interventions that also demonstrated short term effectiveness. Two answers may be posited. First, whilst drawing on evidence from a number of US studies, there existed an English study (rated '+') (McCambridge and Strang, 2004; 2005) from which it was perceived that results could be unproblematically applied in the English context. Second, as is acknowledged in the recommendations (National Institute for Health and Clinical Excellence, 2007), one of the co-authors of the English study was able to contribute (outside of the formal committee meetings) to the refinement of the policy recommendations by providing additional information and guidance. In a similar vein to the earlier discussion regarding the Coggans et al (2002) critique of the LST programme, evidence does not speak for itself; rather, it requires a professional network to frame it appropriately in order to inform the policy making process.

The committee meetings regarding the substance misuse review were a fraught process. It cannot be expected for such meetings to progress rapidly and smoothly to a consensus, especially in complex areas of social policy where a multi-disciplinary committee is required. However, what marked out the substance misuse committee meetings was not so much the lack of consensus, as the lack of disagreement. In a multi-disciplinary group it might be expected that professional ideologies would clash. However, in this instance, the majority of the committee members struggled to find the evidence that they felt was needed to start drafting recommendations that would be applicable to the groups identified in the scope of the review. This was despite the production of further analyses of the data by the RT for the first committee meeting in September, the strong efforts by the MT to prompt the committee to consider particularly promising interventions (RD 2636-2644, 2760-2761, 2811-2814 and 3213-3216 04/09/06; RD 3624-3634 13/10/06), and the collaborative efforts made by the managing and review teams to produce further extensive analyses of the data (in the form of single-study forest plots) for the October meeting. Arguably, this points towards there being significant difficulties in applying the review methods used in this SR to synthesize evidence to inform this area of social policy. Attaining the necessary level of abstraction from such a large and diverse evidence-base proved to be highly problematic, not least because of the lack of established methods for doing so.

Furthermore, the committee were largely reliant upon dominant discourses surrounding substance misuse. This was not because of a conscious commitment to a particular ideological position, but simply because the committee converged around the treatment of 'the substance misuse problem' within a discourse based upon the moral pathological model and public health. This discourse positions the (potential) substance user as being unable to exercise control over their use of substances; the logical public health response, in the same manner in which (for example) sanitation largely prevents the transmission of infection in waste, is to identify the most effective intervention by

which to prevent vulnerable and disadvantaged youths from ever using substances[90]. This discourse was furthermore bounded by an Evidence-Based Policy and Practice (EBP) discourse in which identifying 'effective interventions', rather than exploring the causal mechanisms between the environment and substance misuse, was the driving force for the committee's deliberations. The efforts of the co-optees to draw attention to the importance of the wider determinants of health for deliberating upon the issue were, for the most part, prevented from progressing any further by dint of the fact that they could not be accounted for and discussed within the framework of the dominant discourse.

Finally, it is worth noting that despite the apparent suitability of a SR method that prioritizes the internal validity of studies to produce a review for a committee that is endeavouring to identify effective interventions, in this instance the method did not produce a review that adequately served the purposes of the committee. By no means did the committee accept uncritically the evidence produced by RCTs. Even though the committee were strongly cognizant of the role and importance of RCTs, they found the utilization of RCT findings in the drafting of recommendations highly problematic, both because of doubts concerning their external validity and the awareness that the wider determinants of health were not being acknowledged adequately. This analysis challenges the assumption that evidence from RCTs is utilized uncritically in the formation of policy at the expense of evidence produced using other research methods (for example, see Holmes et al., 2006; Rycroft-Malone, 2006). The substance misuse committee strongly wanted different forms of evidence in order to inform their deliberations, but they were frustrated in their ability to access this by the review methods and the process surrounding the drafting of recommendations.

[90] This 'prevention' approach has a much longer and more involved history than is suggested in this public health discourse, the genesis of the approach having been formed substantively through ideas about the control of marginalized groups and the heavy influence of the US upon drug policy (see Chapter 3).

Summary

Utilizing SR methods that prioritize assessment of internal validity results in an evidence-base that is dominated by evaluations of interventions at the level of the individual or family, rather than the wider determinants of health. Policy makers are aware of this, but in this case study they struggled within the bounds of an 'evidence-based' discourse to argue for a more inclusive SR. The lack of a framework in the SR regarding external validity resulted in the policy network drawing substantially on the findings of just one United Kingdom (UK) study, rather than making considered inferences from a larger body of evidence.

The phenomenon of substance misuse was deliberated by the committee using a public health discourse that positioned substance users as deserving of help. This positioning foreclosed a number of other options; for example, a penal approach, or one where currently illicit substances are regulated. In this case study, a public health discourse interacted with an EBP discourse in such a way as to mean that the identification of interventions at the level of the individual were prioritized over the exploration of causal mechanisms between the environment and substance misuse. Committee members were arguably aware of this, as the also were of the significance of the way in which they framed the issue, but the boundaries of the dominant discourse inhibited them from further exploring these issues.

Chapter 7

Campbell and Discourse: The Development of Systematic Review Methods for Social Policy

The analysis and critique of the systematic review (SR) on interventions to decrease the misuse of substances by vulnerable and disadvantaged youth (presented in Chapters 5 and 6) would simply remain a critique unless its arguments can be taken forward in a constructive fashion. This chapter therefore seeks to increase understanding of how SR methods may be developed in order that social policy making may be better informed. The central argument of the chapter is that the concepts of internal, external, and construct validity (see Chapter 2, sections 1.1 and 3.1) are under-utilized in efforts to develop SR methods for social policy. Furthermore, and of crucial importance, it is argued that paying closer attention to discourses about SR methods and (in this case study) substance misuse would substantively facilitate the development of SR methods for social policy. If this analysis is at least partially correct, then the starting point for developing SR methods should be to pay critical attention to how SR methods and substance misuse are conceptualized and discussed. In this way, a better understanding may be reached regarding how these conceptualizations substantively affect the methods that are advocated and the knowledge that is produced.

This chapter presents an analysis of interview data with members of the managing and review teams in the case study in order to do three things. First, their analysis of the strengths and weaknesses of the SR review method used will be compared and contrasted with the critique presented in this thesis. Second, discourses about SR methods will be considered for the insights that they can give into how the development of methods is conceptualized. Third, and with specific reference to construct validity, discourses about substance misuse are considered; how might a SR

rooted in a different discourse have produced a different body of knowledge? Would this body of knowledge be able to substantively inform policy and practice?

1.0 Practical issues

A substantial number of practical issues were identified as impacting upon the outcome of the SR in this case study, few if any of which are generally acknowledged in the literature on SR methods. It might be argued that the institution undertaking the SR in this case study has unusually tight deadlines, but it should also be borne in mind that SRs are conducted in order to inform pressing policy and practice issues; completion within a timeframe measured in months rather than years therefore tends to be inherent to the task.

The volume of research that needed to be appraised and synthesized by the review team (RT) within the time available to them was identified as placing significant limitations upon the team's ability to fully implement 'systematic' review methods; it also limited their ability to fully critique and synthesize all of the evidence (Chapter 5, sections 2.1 and 3.0). The view of senior team members upon this was that whilst there were indeed risks associated with conducting large reviews rapidly, these were outweighed by the timely delivery of knowledge that could be utilized by a committee to formulate recommendations. This view was grounded firmly in an understanding of the committee's role of *interpreting* the (always contestable) findings of a review;

> "... if all the evidence is ever going to tell you is the direction of travel, being more precise about what the evidence says is looking backwards to the evidence rather than forwards to the recommendation." (ID6 786-789)

In this view, the expectation that a SR in public health can unequivocally *direct* policy is misguided, and the difficult balance to be struck is synthesizing evidence at the correct level of abstraction. This level would maintain the integrity of the findings whilst permitting considered inferences to be made.

The analysis by the managing and review teams regarding the impacts of the timeframe upon the review contrast strongly. The RT expressed substantive concerns about the manner in which the methods focused their efforts upon data extraction rather than data synthesis (Chapter 5, sections 2.1 and 4.1). Whilst the managing team (MT) sympathized with the intense stress experienced by the RT in completing the review on schedule, their perception of this stress sharply contrasted with that of the RT's, where the stress was bordering on (and sometimes crossing into) the unmanageable (Chapter 5, section 4.2). The conduct of a comprehensive SR is intended to be a means of attaining an unbiased grasp upon a mass of evidence (Egger et al., 2001a; Chalmers, 2003; Petticrew and Roberts, 2006), but in this case study the methods utilized threatened to engulf the RT in pieces of evidence rather than facilitate their endeavours to synthesize it. The analysis presented here suggests first that there is a substantive difference in the perception of stress between those tasked with managing and those actually conducting the review; moreover, this substantive difference extends to the perceived impact upon the production of rigorous knowledge by the review. The MT proposed that the time constraints did not impact deleteriously upon the syntheses produced for the committee, an analysis that the RT's interview indicated that they were less sure of.

The SR methods literature is also near silent upon how, if evidence does indeed require (deliberative and potentially time-consuming) interpretation in order to rigorously make policy, a committee might cope with a quantity of evidence such as that presented in the review in this case study[91]. The committee in this case study experienced significant difficulties in deliberating fully upon the eighty-six evidence statements produced by the review, both in terms of the time available to discuss them (Chapter 6, section 3.0) and in attaining and maintaining a grip upon them in order that the evidence was utilized in the committee's deliberations (Chapter 6, section 4.0).

[91] Although it should be acknowledged here that Bayesian approaches have been advocated as a transparent means of modelling decisions based upon a combination of extensive quantities of prior distributions of effects (the 'expert view') with likelihoods (trial data). A balanced overview of Bayesian approaches is provided by Pope et al (2007, p.55-67).

These difficulties were acknowledged in interviews with senior team members, and the proposed solutions tentatively couched in the terms of an ongoing debate over how best to synthesize and present evidence to a committee (ID6 505-530; ID7 508-539).

2.0 Expertise: Essential or a source of bias?

'Expert opinion' is placed at the bottom of the traditional hierarchy of evidence on the basis that it represents a non-systematic, opaque interpretation of the evidence. Expertise in this formulation, then, is only to be utilized if other, more rigorous and less biased, forms of evidence are not available. Members of the review and managing teams had considerable difficulty accounting for their use of expertise in the conduct of the review; whilst their experience was that the prior knowledge of researchers and practitioners was vital for conducting an informed review, they could not justify utilizing this knowledge at the same time as advancing the argument that SRs represented a substantive departure (with regard to transparency) from the 'unsystematic, narrative' reviews of old (Chapter 5, section 2.1).

Whilst a catholic approach to evidence is nowadays generally considered to be desirable, it is perhaps not surprising that the teams experienced substantial difficulties in balancing and integrating the different forms of knowledge in the review. In the wider field of Evidence-Based Policy and Practice (EBP), the development of SR methods progressed, meaning that there is now a greater depth and breadth of work to draw upon to inform the selection of review method (Popay et al., 2006; Gough, 2007; Greenhalgh et al., 2007; Pope et al., 2007). The view expressed by a senior member of the MT in this case study, and which it was stated also had the broad support of the relevant committee, was that an expert summary of the state of knowledge in the field to be reviewed was very important both for constructing the review's scope and as a starting point for the committee before starting to deliberate the evidence (ID6 338-369). Whilst this proposal does not formulate precisely *how* this expertise is to be integrated systematically and transparently into a review, it does represent a strong

acknowledgement that expertise is an essential part of a review that cannot be properly appreciated using the hierarchy of evidence alone.

The proposal described above regarding a 'non-systematic, narrative' review that concisely summarizes the state of knowledge in an area, both with respect to current policy and practice and the philosophies that inform different approaches, might also address some of the concerns expressed by the RT with regard to the integration of stakeholder knowledge. The RT expressed the view that their task was, in certain respects, made more difficult by the scope changing to reflect stakeholder input rather than building upon what they regarded as the imperfect, but stronger, research base that already existed (Chapter 5, section 2.2). It is unlikely in any area of social policy for there to be complete congruence between (or even within) groups of practitioners, researchers, and policy makers, so it is to be expected that differences would arise regarding the scope for a SR on substance misuse. However, it is notable that the RT (who were present at the stakeholder meeting) felt marginalized from this consultative process, believing that they could have provided substantive input[92]. Whether or not the scope would have been improved through greater input by the RT is an empirical question that this research cannot answer. It is notable, however, that outside of instances where a clear consensus is attained (see ID6 126-159) the methods for integrating knowledge from researcher and practitioner perspectives is only partly formulated.

The case study in this research demonstrated how, even if methods for the integration of expert knowledge into SRs are unformulated, this integration nevertheless occurs (Chapter 6, section 2.1). The pressing issue here is how this 'unsystematically' generated knowledge may be 'systematically' integrated into a

[92] There is an unresolved issue here surrounding the role that 'expertise' (whether that of an individual academic or practitioner, a RT, or a committee) should play in a SR. Upon what basis the RT believed that their expertise could inform the review's scope whilst maintaining their impartiality in the review process is unclear and is not explored further here. Further analyses (in addition to this chapter) of the ambivalence that exists over the role of expertise in SRs are given in Chapter 5 (section 2.2) and Chapter 6 (section 4.0).

review. In short, it cannot be integrated 'systematically' using the methods for conducting a 'systematic review' as set out by the managing institution. This is because the knowledge concerned is not in the form of a report of primary research that can be critically appraised on the basis of its internal validity; rather, it is a critique of research of substantial import for the SR concerned, but which requires critical appraisal on a different basis. The fact that the review and managing teams decided to include this knowledge was pragmatic and justifiable on the grounds of utilizing relevant expertise, but doing so involved stepping outside of the SR methods protocol. It is this protocol that is maintained to play a key role in producing knowledge that is more rigorous and less biased, and 'breaking' the protocol in this manner, if adopting a strictly 'evidence-based' approach as would be expected in the institution concerned, raises questions about the neutrality of this section of the review.

3.0 Research validity

In Chapter 2 (section 3.1), the inherent trade-offs that occur in both primary research and SRs between the different forms of research validity (internal, external, and construct) was discussed. In this section, the emphasis given to each form of validity in the SR analysed in this case study is considered with regard to the effect that this had upon the production of an evidence-base.

3.1 Internal validity

In the interviews conducted with review and managing team members in the course of this research, the role of internal validity was not explicitly raised (either by myself or the interviewees) with regard to the development of SR methods. It will, however, be argued here that the extent to which a primary role was ascribed to internal validity in the conduct of the SR in this case study limited its potential to produce knowledge that was usable by a committee in its deliberations. These limitations occurred in three ways; in conducting the review in a manner that was attempting to assure its own internal validity, in the focus upon internal validity as the

primary arbiter of study quality, and in the degree of transparency that was attainable using these methods.

The internal validity of the SR in this case study was understood by the review and managing team members to hinge upon it being conducted according to the methods protocol. In this way, it is argued, the committee would be able to have more trust in the rigour of the review's findings and hence make policy that was explicitly based upon the best, most unbiased, synthesis of evidence available. The conduct of a SR, especially one that needs to be delivered within a tight timescale, requires that a balance be struck regarding the allocation of time to different tasks; in short, what is the opportunity cost of (for example) adhering closely to a protocol that emphasizes internal validity as against (for example) using the protocol as a guide that may be interpreted with care? In concrete terms, what was the opportunity cost of spending considerable quantities of time in the review recording inclusion/ exclusion decisions, all of the studies' data, and the debates over exactly what 'type' an individual study was (in order to place it on the hierarchy of evidence) (Chapter 5, section 2.1), as against a more flexible interpretation of the protocol? Again, this is an empirical question that this research cannot answer, but which is of substantial importance.

In focusing upon grading studies primarily upon internal validity criteria, the SR in this case study produced some evidence statements that the review and managing teams felt required additional explanation if they were not to be misinterpreted. In Chapter 6 (section 2.1), evidence statement 21.1 is highlighted for the support it offers the Life Skills Training (LST) programme on the basis of the internal validity of the studies evaluating it. However, the additional critique of the programme (Coggans et al., 2002) that the teams argued it was necessary to be cognizant of if informed policy was to be made was not in the evidence statement itself, nor could it be put into the evidence statement as it was not a primary study that could be appraised in the manner set out in the protocol. This analysis may appear to be overly critical; after all, if

a strong case was made for the critique of the LST programme to be included, what does it matter if it was not formally found on a database, appraised, and included according to the protocol? Arguably, however, this issue *is* of substantive importance. This is not only because the review should, in order to be comprehensive, have searched for other critiques on this and other programmes, but also because it is not clear whether evaluations of the LST programme in later studies take account of Coggans et al's critique. This is of crucial importance, for if one can only make considered judgements (even if only *some* of the time) regarding the quality and rigour of studies against a background of the debate in a particular field, then the role to be played by generic assessments of internal validity decreases considerably. The use of Coggans et al's critique in the SR in this case study indicates that this was tacitly understood by the review and managing teams, but in view of the fact that its use fell outside of the methods protocol, they felt unable to justify its use.

Finally, the issue of utilizing knowledge that was generated by a means outside of the protocol raises substantive questions about the transparency of the review. If 'deviation' from the protocol is considered to introduce bias into the review, then clearly the inclusion of Coggans et al's critique introduces bias; it is perhaps not surprising that the role the critique played in guiding the committee's deliberations (Chapter 6, section 2.1) is not highlighted. The issue here, however, is the degree to which an observer of the SR in this case study could critically analyse its findings. The methods protocol is intended to make the SR process as transparent as possible. However, in this case study it was necessary for the review to diverge from the protocol in order to include substantive and relevant knowledge, but without including any details of how this knowledge was found or utilized. In this sense, the focus upon the assessment of the internal validity of studies as a means of assuring rigour and transparency in a SR is some way wide of the mark.

3.2 External validity

The committee in this case study experienced considerable difficulty in utilizing the review as a basis for making inferences from the evidence in order to make recommendations for policy (Chapter 6, section 4.0). In analysing this process, I have made the case for there having been substantive evidence in a number of the studies *included* in the review that (adhering to the methods protocol) was not extracted; instead, the focus was upon stating with greater accuracy what interventions were effective (in the environment in which the studies were conducted) rather than providing evidence that could inform the committee in making inferences to the areas with which their policy making was concerned. For example, the socio-economic conditions in which research subjects lived in the United States of America (US) were posited to have substantial similarities with that of disadvantaged and vulnerable youth in the United Kingdom (UK), and the nature of one study's findings (where secondary outcomes were robust across a wide range of research subjects), were both argued to have had the potential to inform the committee's deliberations if the review had extracted the relevant knowledge from them. The focus upon the effectiveness of interventions also directed attention away from what was arguably vitally important information regarding the large differences in resources required (an essential component of the economic modelling regarding cost-effectiveness) to deliver an intervention (Chapter 5, section 2.2).

In the analysis of the SR review in this case study, the case was also made that evidence from a number of *excluded* papers could have substantively informed the committee's deliberations (Chapter 5, section 3.0). This argument, made with the benefit of hindsight following the committee meetings, questions again whether a methods protocol that focuses upon internal validity as the primary criterion upon which to appraise studies adequately serves the needs of a committee concerned with making recommendations for social policy. Knowledge regarding an evaluation of

'culturally-specific' interventions addressing substance misuse (Chapter 5, section 3.1) and the importance of focusing on interventions that are robust when delivered under conditions some distance from the ideal (Chapter 5, section 3.2) was excluded from the review on the basis that the papers concerned were 'non-systematic' reviews. Moreover, the identification in studies excluded from the review regarding the benefits of considering determinants of health at a level wider than those focused upon when delivering an intervention to an individual, did not meet the criteria for knowledge that was admissible for the review (Chapter 5, section 3.2).

As with internal validity, there was little discussion in the interviews conducted for this case study with regard to external validity. One senior team member commented upon the difficulty of striking a balance between making very specific recommendations that were not very widely applicable, and making recommendations that were widely applicable but which were at such a level of generality that they added little if nothing to the evidence-base (ID7 546-550 and 558-561). Within the wider SR literature, there is also a growing acknowledgement of the role that explicitly considering external validity can play in evidence-based policy and practice (Boaz and Pawson, 2005; Saltz, 2005; Bonell et al., 2006; Arai et al., 2007; Glasgow et al., 2007). Whilst this growth acknowledges that external validity is important, difficulties remain over how to integrate it with the focus upon internal validity that is one of the dominant features of SRs as currently conceived and conducted. The following section therefore considers the manner in which the issues surrounding external validity that arose in this case study could be addressed.

The essence of all the issues around external validity that arose, with the exception of the lack of details regarding the resources required to deliver certain interventions (Chapter 5, section 2.2), was that there was not a framework within which inferences could be made. For example, on what basis could findings from a study of youths from 'working poor' African American families be utilized in UK policy making?

Are correlates of youth substance misuse (for example, repeated truancy) utilized in a US study sufficiently similar to those in the UK to allow findings to be generalized? Is the identification in papers drawing upon other research conducted in the US of a particular approach to 'culturally-specific' interventions applicable to the UK, or is the nature of migrants' integration into British society substantively different to that in the US? Is the identification of wider environmental contributions to youth substance misuse in the US (such as poverty or neglected public services) also applicable to the UK? All of these questions highlight the manner in which the review did *not* make a case for utilizing a theoretical framework in order to guide both the structure of the review and the inferences of the committee. By default, the review and the committee adopted an unspoken and mixed framework. In the examples given above, of the evidence that was included in the review, that from the US was rated as 'C' (of uncertain applicability to the UK). The point here is that these studies and their potential to inform UK policy were categorized on the basis of their geographical origin, rather than upon any other criteria. Doing so demonstrates the use of a theoretical framework (one in which social conditions and subsequently substance misuse in the US differ substantively from those in the UK) as much as if an explicit theoretical framework (where, for example, phenomena x, y, and z were posited to drive substance misuse) had been utilized.

3.3 Construct validity

There is a fine distinction to be made between external and construct validity (Chapter 2, section 1.1). This accounts for the thrust of the analysis presented here bearing similarities with that presented above regarding external validity. Most pertinently, this is with regard to the utilization of a theoretical framework in order to guide conceptualizations of the phenomena being reviewed and the subsequent interpretation of the evidence in order to make recommendations for policy. Upon initial examination, the suggestion that is made using the concept of construct validity that there exist different theoretical frameworks that may have equal validity for the analysis

of complex social phenomena might be thought to be highly controversial in the evidence-based field. However, the interviews conducted with senior team members highlighted the awareness at this level of the importance of being clear about the theoretical framework used in the review (ID6 364-369), how constructs might actually apply across a wide range of health behaviours (ID6 186-209), and of how the utilization of different constructs could substantively alter the evidence that is sought and utilized in the review (ID7 709-735).

The analysis of the review process in this case study investigated the way in which constructs that were advocated in some of the papers excluded from the review might have provided a useful theoretical framework (Chapter 5, section 3.3). These papers advanced constructs regarding substance misuse by youths that differed strongly from those that were utilized in the review. For example, constructs of the process of deciding to use substances as being 'rational' (as opposed to emotive), or of the introduction to substance use being a passive (as opposed to actively negotiated) process, were questioned. In a similar way, the basis for classifying certain substances as illicit (and therefore requiring a response couched in penal terms) or for analysing substance use in terms of cultural reproduction rather than the moral pathological model[93], fell outside of the discourses that guided the conduct of the review. This meant that potentially important constructs were effectively frozen out from consideration.

The point of summarizing these alternative constructs here is not to argue that they are necessarily better than those utilized in the review in this case study; rather, it is to highlight the importance of being explicit about the rationale for using particular constructs in a review. Arguably, the process of formulating, completing, and deliberating upon the SR in this case study took place without an awareness of the substantive effect that the framing of the issue had upon the evidence that was

[93] Greater detail is provided regarding these alternative constructs in the conclusion to Chapter 3.

selected and the manner in which it was utilized in the formulation of policy recommendations. The argument in short is that a moral pathological model of addiction that is rooted in nineteenth century evidence of dubious provenance has become so thoroughly incorporated into the narratives that inform policy making that the policy network has become totally unaware of its role in framing drug policy. The implication of this is that the policy network would benefit from a greater awareness of the genealogy of how substance use has come to be framed in the UK. This would require awareness of not only the roots of the moral pathological model, but also the substantive impact that the wider political-economic environment, attitudes towards ethnic minorities, women, and the working class, and the aspirations and machinations of professional groups have had upon the risks attributed to substance use and the formulation of policies that seek to regulate substances (see Chapter 3, sections 2.0 and 3.0).

4.0 Researching substance use in the young: Issues covered in papers that were excluded from the review

An important part of the process of developing a theoretical model for the conduct of a SR in social policy (sections 3.2 and 3.3 of this chapter) is the balancing of contentious issues in the field concerned. Deliberating upon the evidence and taking a position on these issues at the outset, a situation that is quite distinct from being 'biased' in some way in conducting the review, is arguably of vital importance in two ways. First, it clarifies the rationale for taking a particular position rather than obfuscating it behind a label of 'objectivity'; and second, it should enable the review to be better based upon the cumulated learning in the field.

In the analysis of the review in this case study, a number of contentious areas were identified (in excluded papers) that might have usefully informed the review's search strategy, study inclusion criteria, and the interpretations of the evidence by the committee. First, the inherent problem of retaining research subjects over extended

periods of time has led some researchers to argue that applying conventional measures of rigour (as required in randomized-controlled trials (RCTs)) is unjustified (Chapter 5, section 3.2). Second, the focus upon discrete interventions has been claimed to miss out on an understanding of the impacts of (for example) the wider school environment upon a range of behaviours, of which the misuse of substances is just one part (Chapter 5, section 3.2). Third, expanding the focus from interventions in a programme to the socio-economic environment in which substance misuse takes place may produce some quite different understandings of what drives problematic substance use and how it might be better addressed (Chapter 5, section 3.3).

It is not claimed here that attending to these areas of debate would have revolutionized the conduct of the review, nor that they necessarily represent *the* most important areas that should have informed the conduct of the review. It is argued, however, that in the pursuit of precise knowledge solely about the 'effectiveness' of interventions demonstrated in studies, the review excluded knowledge of substantive importance. This knowledge is argued here to have had the potential to inform the conduct of the review and the deliberation of the evidence to an important degree; dispensing with the need of engaging with the knowledge (in the form of debates) in this area confers an illusory level of clarity to the evidence synthesized in the review. If it is acknowledged that it is necessary to simplify complex reality in a review in order to facilitate decision-making, the question remains as to what level of simplification is warranted.

5.0 Knowledge synthesis

The problems encountered by the RT in synthesizing evidence were substantial. The difficulties hinged around whether the interventions and/or the research subjects were sufficiently similar to be able to justify cumulating the outcomes data. The heterogeneity of the data in this review ruled out performing a meta-analysis; the MT decided upon the presentation of data to the committee in the form of 'forest

plots' for individual studies as a means of facilitating the interpretation of the evidence (Chapter 5, section 2.0 and Chapter 6, section 3.0). In the event, the committee made little use of these 'forest plots', with some members instead pressing for evidence that would give them a better understanding of the causal mechanisms at play (Chapter 6, section 4.0).

The difficulties involved with synthesizing data were acknowledged at a senior level within the MT to indicate the need for the substantive development of methods. This development questions core assumptions within knowledge synthesis for social policy as set out in the methods protocol used in this review. This certainly does not mean that *cumulative* knowledge synthesis (in the sense of pooling effect sizes so as to attain greater statistical power) was in some way wrong, for:

> "... when you look at all those forest plots [in clinical drug trials]... [they] are all relating to the same thing... and you can overturn some of the great nostrums of clinical medicine which were doing more harm than good." (ID6 430-436)

Whether it is possible to rigorously pool results in the same way in public health interventions (whether widely or narrowly defined) is less clear given the variability in the delivery of interventions when compared to, for example, the administration of a medicinal drug (ID6 436-444). SR methods in public health were posited to require something more than a solely cumulative approach to evidence synthesis:

> "...our challenge is to try to find the method that will allow us to review large quantities of data, but not synthesize it in the building block kind of way, but rather to think about, given that we've got this mass of stuff, how can we find our way quickly to the things that we need to take account of, but to do that in a way that doesn't introduce yet another layer of bias. And we haven't cracked that yet." (ID6 451-456)

6.0 The development of systematic review methods for social policy: Insights from this case study

This section endeavours to take a step further on from the critique provided above of the process of conducting the SR in this case study. In taking this step, the aim is to get nearer to identifying ways in which SR methods for social policy could be

improved. The analysis presented here draws upon the interviews conducted with members of the review and managing teams in order to present an 'insider' (emic) perspective on the development of methods. Such analyses are conventionally presented using the respondents' own words in an effort to preserve the authenticity of what is said and to minimize the interpretation involved in the analysis; given that the majority of the respondents in this case study did not give permission for direct quotations from their interviews to be used, the analysis presented here has necessarily relied upon the considered interpretation and summary of responses. The significance of presenting this analysis is the manner in which it allows the diversity of respondents' thought on SR methods to be expressed. Whilst no clear methodological solution is provided by the analysis, it does allow ambivalence over the merits of traditional methods to be expressed. Respondents sought to defend what they felt to be the strengths of the traditional approach, whilst also attempting to express ways of developing these methods. It was often difficult for respondents to conceptualize their methodological proposals within the dominant discourse of EBP and traditional SR methods. It is this difficulty that provides a central insight; for SR methods in social policy to develop, close attention needs to be paid to the way that research methods and the topics of enquiry are conceptualized.

In designing the presentation of the analysis in this section, consideration was given to clearly defining the role played within the review and management teams by each of the respondents. Aside from the risk that respondent anonymity would not be preserved if such role identification was utilized, in the final analysis there is arguably little to be gained by identifying these roles. Whilst differences in power between the management and review team clearly exist, it is not the case that senior members are dictating a particular approach. Indeed, discussion about the development of methods by all members of the RT was valued by the institution concerned. Distinguishing between respondents on the basis of their position would therefore serve no purpose. However, it can be noted from the analyses presented here that the more senior the

respondent, the more developed is the analysis of SR methods. This is perhaps unsurprising given the role at management level of reflecting critically upon the methods used across a range of reviews in public health. However, this is in itself significant, for it demonstrates a commitment and openness to methodological development at a senior level.

6.1 The focus on effectiveness

Respondents in the case study were clear that the crux of what the review was intended to do was to establish what interventions were, or were not, effective in reducing the use of substances by vulnerable and disadvantaged youth. However, respondents differed with regard to the methods advocated for establishing this effectiveness. For some, the traditional SR methods were of primary importance given the crucial role they played in extracting effectiveness data into a database that subsequently allowed the objective analysis and synthesis of that data (ID3 525-547). Indeed, one respondent felt that the review could have been made stronger if this data had been recorded in a more fundamental form that would have better allowed statistical manipulation, rather than working with interpretations of that data (ID3 557-583).

Adhering to the requirement of establishing effectiveness was generally acknowledged to have limited the review with regard to extracting and synthesizing data that would inform implementation of the interventions. This limitation was viewed first as resulting from fairly minor methodological limitations relating to the format of the evidence tables and evidence statements (ID3 777-782; 960-963); if these were re-engineered so as to require evidence on implementation issues, it was argued that it would then be possible to conduct the review in such a way as to provide this evidence. However, there was also acknowledgement of the tensions in the review between comprehensiveness and a more limited, but potentially more usable, extraction of data within the resources available. For example, the vast quantity of evidence that a

comprehensive approach generates was viewed as prohibiting critical consideration of questions other than those relating to effectiveness (ID2 438-445). For others, this was not such an issue, as the primary aim of the review was to determine what interventions were effective. Issues surrounding implementation were viewed as important, but secondary (ID1 145-147; 167-175).

Other respondents proposed more fundamental methodological developments. The proposals are significant in view of the 'lack' of any great similarity between them; this does not mean that the proposals are irreconcilable, rather it can be taken as a healthy sign that there is substantive thought behind the proposals and that active discussion over their relative merits can take place. First, it was proposed that in order to inform the critical analysis of data, a key aspect of a review should be an understanding of the causal pathways of a phenomenon (ID6 585-601). For example, there may be pathways that can be identified regarding the initiation and continuation of substance use, and pathways by which an intervention achieves its effectiveness. Another perspective on methodological development was that the emphasis in the review upon stating effectiveness (in the form of effect sizes) as accurately as possible was misguided in the field of public health, where there exists a significant risk that the lack of control over experimental conditions could make such statistical accuracy spurious. For this respondent, a far broader approach was necessary; on the basis that there already exist well-developed understandings of what interventions are effective in public health, the key issue is to determine what is *cost-effective* (ID7 46-71; 187-212; 233-257; 582-599).

6.2 Evidence synthesis

Respondents' analyses of the process of evidence synthesis were notable for the degree of interpretation that was argued to be inherent to the processes of data extraction and synthesis. This is surprising given the emphasis in the traditional approach to SRs to the objective cumulation of evidence by a technical process, such

as in a meta-analysis[94]. In this case study, it was not the case that respondents were simply arguing that evidence synthesis was a subjective process, for the methods utilized were mostly defended as being objective and transparent to the greatest degree possible. However, it was proposed that subjectivity was an inherent part of the process that an adherence to the technical processes embodied in methods could only partially resolve.

Elements of subjectivity were suggested to operate at most stages of the SR process despite adherence to the methods protocol. In view of the large amount of data that many pieces of research produce, this could occur with regard to the selection of the data that is extracted and summarized from a paper (ID1 299-310; 455-466). It could also occur in a more fundamental way, in that the review in this case study captured reality in a certain way (predominantly through academic research where outcomes were quantified) at the expense of other substantive knowledge produced by other means (ID5 459-486). However, not all respondents would concur with this view. It was also argued that data extraction is mechanical and transparent, but in view of the fact that the categorization and synthesis of that data is an intellectual process there is the potential for subjectivity to intrude into the process (ID2 515-525).

The necessity in the SR in this case study of relying upon consensus within the scientific community in order to model cost-effectiveness was acknowledged. This was candidly described in terms of the *assumptions* that were made regarding the likelihood of an experimental substance user progressing to chronic addictive use and the economic sequelae thereof. It was necessary to make these assumptions in view of the paucity of economic data relating directly to substance use (ID7 388-400).

[94] Whilst it should not be inferred that proponents of meta-analysis are ignorant of the importance of *interpreting* results, key SR manuals and texts are notable for the absence of discussion about how rigorous interpretation might take place (for example, Khan et al, 2002; Higgins and Green, 2005; Petticrew & Roberts, 2006, p.192-209).

The views of respondents differed in respect of how evidence should be synthesized. One respondent argued that if rigorous evaluations of interventions in the field of substance use were to be made, then long-term RCTs extending from childhood through to adulthood are required (ID7 75-91). Another perspective proposed that evidence synthesis in public health should work with the findings of research that already exists, and that the challenge is to develop SR methods to better utilize this imperfect but substantial evidence-base. In this view, public health interventions are conceptualized as complex, rarely being discrete and replicable in the sense of a clinical drug trial (ID2 129-134; 608-626; ID6 430-442). The implication for public health is that evidence synthesis is more likely to require a process of knowledge interrogation rather than the cumulation of effect sizes as in a meta-analysis (ID6 437-456).

6.3 Adherence to a methods protocol

The rationale for adhering to the methods protocol in order to produce a rigorous and defensible SR was not subscribed to by all of the respondents in this case study. On the one hand, adherence to the protocol was used as a way of explaining why the review was produced in the way that is was; the protocol prioritized identifying what interventions were effective, therefore the hierarchy of evidence was utilized and the review focused upon RCTs (ID3 955-963). On the other hand, the view was expressed that keeping strongly to the protocol might actually inhibit the conduct of the review if there was evidence that adopting a more flexible approach would ultimately facilitate the production of an evidence-base that would be of greater utility (ID6 372-381). This view was encapsulated with respect to reviewers attaining the correct level of detail in their data extraction and synthesis; doing so is not a precise science, there instead being a "real art" to the process (ID6 489-496). This suggests a substantive role for reviewers with subject-specific expertise, rather than the reliance upon a methods protocol to assure rigour.

6.4 The utilization of expertise

There was considerable ambivalence expressed by the respondents in this case study regarding the role that expertise played in the conduct of a SR and the making of recommendations. The exclusion of 'non-systematic reviews' from the substance misuse review was justified on the grounds that such reviews first place an over-reliance upon chance to identify relevant evidence. Second, the interpretation by the authors (whether 'expert' or not) of the included evidence was not considered to be transparent (ID2 360-366). It was argued that the potential for methodological development existed, in that these 'non-systematic reviews' could potentially be included in a SR, but upon the basis that they would need to be critically appraised using the same criteria as 'systematic reviews' (ID2 307-316). Nevertheless, the expertise of reviewers in producing a rigorous and defensible SR, and of the committee in interpreting all of the evidence placed before them was acknowledged be of substantial significance (ID1 213-227; ID2 339-348; 641-664). Whilst efforts were made by these respondents to distinguish between an expert 'expressing an opinion' (whether in print or at a committee meeting) and 'interpreting the evidence', it was not clear exactly how these were proposed to differ aside from the fact that the latter took place in the context of a SR.

A more explicit belief in the role of expertise was expressed by other respondents. First, the importance of subject-specific expertise in devising the scope for the SR was highlighted; to overlook this risked the non-utilization of a history of cumulative learning in the field (ID3 264-286). Second, methodological development was proposed in the form of revising the traditional approach of comprehensively reviewing the literature; basing search methods upon an explicitly defined model of human behaviour was advanced as a means of enabling a more "forensic" approach that could better inform the committee in drawing up recommendations (ID6 803-813). Finally, the role that the expertise of the committee played was clarified by one respondent in setting out how the evidence-base (in the form of a SR) could inform

their deliberations. It was proposed that the evidence does not direct the committee as such; "human thought, logic, knowledge, and expertise" has to be applied to the evidence in order to draw up recommendations (ID6 509-551).

6.5 Making inferences to the UK context

The one area of the SR in this case study about which consensus existed was with regard to the issue of making inferences from the evidence to the context of the delivery of services in the UK. It was openly acknowledged that the framework for assessing applicability to the UK context was not well developed and that it was left to the committee to make inferences in this respect (ID1 399-413; ID2 343-348; 689-697). However, there was some disagreement over where responsibility lay for assessing how applicable the results of studies were to the UK. The limited framework for assessing applicability set out in the methods protocol meant that the team conducting the SR itself (in the context of a very large volume of studies to synthesize in a limited timescale) did not schedule time for this task, instead viewing their primary role as synthesizing the evidence of effectiveness (ID3 881-887). The task of making inferences to the UK context thus fell to the committee almost by default, based upon the quite crude applicability criteria that had been applied to the evidence statements (ID2 690-697).

Conclusion

In Chapter 2, it was proposed that Donald Campbell's body of work represented an under-utilized resource for thinking about and improving current SR methods. Does this case study provide support for this proposal? The framework of internal, external, and construct validity, and in addition the role of the scientific community, will be used here in order to consider the strength of the proposal made in Chapter 2.

Campbell's advocacy of the experimental approach in evaluating social policy met, not surprisingly given the prominence of this approach in the hierarchy of

evidence as the primary means of obtaining internal validity, with general agreement by the respondents in this case study. However, the extent to which respondents felt that a *judicious* utilization of the experimental approach (given the inherent difficulties of implementing such research rigorously in open systems such as in public health) was notable. Campbell argued that quasi-experimental approaches might be more applicable in these circumstances; the review and managing teams in this case study recognized that a well conducted study from lower in the evidence hierarchy could better inform the review than a poorly conducted RCT, but their reluctance to deviate from the methods protocol meant that internal validity (as ranked in the hierarchy of evidence) was prioritized. Arguably, the methods protocol needs to better accommodate the judicious use of different research methods according to the limitations of the field of study.

It was argued by Campbell that statistical inference alone represented insufficient grounds for making generalizations from one study to another context. None of the respondents in this case study contested this; the expertise of the RT and the committee was viewed as being of significant importance for interpreting the results of statistical analysis and for considering its applicability in other contexts. It was also acknowledged that the methods protocol did not contain an explicit model for making generalizations from one study, or body of studies, to another context. Campbell asserted that making these generalizations inherently involved assumptions about causal mechanisms in the phenomena concerned, and advised that it was theories about these mechanisms that should be tested in a range of contexts in order to increase understanding of them (Campbell and Russo, 1999). To do this would require a significant revision of the process utilized in the SR in this case study; it would require an iterative rather than linear approach, as well as the accommodation of purposive sampling techniques in searching for and utilizing the findings of other studies. However, it is notable that the kernel of this approach is contained in one respondent's proposals for methodological development, where it is argued that a more 'forensic',

rather than comprehensive, approach to searching the literature might be of greater utility. Whether or not an iterative approach could also be integrated into this development is difficult to say; tight timescales and the demanding nature of public consultation could make doing so highly problematic.

In contrast to the acknowledgement by the managing and review teams of the methodological developments that were required with regard to external validity, the issue of construct validity struggled for recognition. For most (but not all) respondents, construct validity was not so much dismissed as simply bypassed. In the interviews conducted for this research, the topic of construct validity was broached by asking the respondents if they felt that there would have been any other valid questions that the review could have posed; could the review topic have been conceptualized in another, equally valid way? In summary, the response was that the questions asked reflected current policy concerns and thinking in the research field. As such, there was not an awareness of the possibility that there might be other valid conceptualizations, or debates about these concepts (see Shadish et al., 2002, p.66-72), which might have usefully informed the conduct of the SR. More fundamentally, Campbell's contention that all scientific knowledge has a qualitative basis is not acknowledged. This qualitative basis is in the sense that knowledge is necessarily rooted in people's understanding of the world, as formed by their personal experiences and professional training (Campbell, 1978, p.191-193). In this way, the respondents and committee members in this case study are arguably unaware of how their own (unexamined) constructs substantively influenced the design and conduct of the review. This issue is not amenable to a straightforward methodological tweak; to address it would require some fundamental reconsideration of the underlying tenets of each review question asked.

Finally, Campbell's notion of 'competitive cross-validation', where the scientific community assures rigour through a social process of critical attentiveness (Campbell,

1986; Campbell and Russo, 1999), was well understood by the respondents in this case study. The SR was perceived to rest heavily upon the self-policing of the scientific community at all stages, from inception through to the drawing up of recommendations. However, this understanding points to a central concern in the conduct of SRs; there is no absolute, objective starting point upon which to base this critical attentiveness[95], as Campbell was only too aware (Campbell, 1988). The prioritization of internal validity within the hierarchy of evidence is not the result of implementing objectively demonstrable proof that studies with the highest internal validity produce evidence of effectiveness that best guides policy making; it is a consensus obtained within a scientific community that shares a common approach to what constitutes rigorous research. A quite different consensus could be reached in a different scientific community, or through discourse that led to a renegotiation of consensus within the currently dominant scientific community. In this respect, the conduct of science (and of SRs) is an inherently social affair that relies upon the negotiation of consensus and an openness to the critique of dominant methods.

Whilst it has been argued here that an EBP discourse constrains the development of SR methods in certain important ways, it is nevertheless the case that SR methods are still an issue that is open to constructive debate. Respondents in this case study exhibited considerable ambivalence in their views and proposals for the best means of conducting SRs. This could be seen as representing contradictions inherent to attempting to conduct 'systematic' reviews within a policy and practice community that simply cannot be led in the manner encapsulated in the phrase 'evidence-based'. Whilst it would be convenient if policy making could be most rigorously and fairly achieved simply by utilizing 'the best evidence', there is simply too much knowledge and expertise contained within the policy and practice community to be led in this straightforward fashion. However, the ambivalence expressed by

[95] Given the theory-laden nature of observations and the manner in which values permeate all research (see Chapter 1), this should not surprise us. The important issue is that if this is openly acknowledged, there is a path upon which constructive debate over methods can take place.

respondents in this case study is arguably evidence that the members of this review and management team are open to debate and discussion, even if their efforts to do so are stymied by the boundaries of the dominant discourse upon SR methods.

This openness to debate is of considerable importance, for it demonstrates a willingness to consider other approaches to conducting SRs whilst keeping what are perceived to be the strengths of the methods as they currently stand. The analysis presented in this research has highlighted a number of considerable difficulties that were encountered in the course of conducting the SR. It was arguably important for the analysis to focus attention on these issues, for to have avoided doing so would have risked producing a sanitized analysis. However, to seize upon these difficulties is to miss out on the extent of thought shown by the interviewees regarding methodological development. This is not to say that problems do not occur; first, team members still struggled at times to explain their proposals for methodological development whilst still trying to meet some of the more exacting (and arguably, quite narrowly defined) standards of the dominant approach to reviews. And second, the diversity in the proposed methodological developments makes it quite clear that respondents' thoughts have not simply been driven in one direction by an adherence to the traditional SR approach. Instead, there is an ongoing debate about methods that can be fostered. A contribution to this debate, based upon insights attained in the conduct of research into this case study, is set out in Chapter 8.

Summary

Whilst internal validity was considered a *sine qua non* of SRs by the respondents in this case study, there is considerable awareness of the importance of a *judicious* application of methods that prioritize internal validity. External validity is acknowledged to be an under-developed aspect of SRs; to give this form of validity due emphasis would require significant changes to SR protocols based upon the traditional

approach. Methods would need to be iterative (rather than linear) and facilitate (rather than limit or obfuscate) judgements made by reviewers.

In contrast, construct validity is an aspect of SRs about which there is little awareness. The manner in which constructs frame the topics investigated in SRs and the subsequent deliberations upon this knowledge, remains largely unperceived by the respondents in this case study (although some senior MT and committee members demonstrated a certain awareness of the issue). The development of SR methods can draw upon the energy of a thoughtful and self-critical reviewer community, but will need to pay close attention to the use of discourse in order that potential avenues for development are not foreclosed.

Chapter 8

Conclusion

This concluding chapter focuses upon two areas. First, the methods utilized in conducting the research for the case study presented in this thesis are critically reflected upon so as to inform the design of future research in this area. The opportunity is also taken to briefly reflect upon the experience of conducting the research under the auspices of a Collaborative Awards in Science and Engineering (CASE) studentship. Second, Campbell's validity framework is considered in the context of the case study systematic review (SR) for the contribution it could make to the development of SR methods for social policy. The idealist and realist ontological positions are examined with regard to how they conceptualize the different forms of research validity; and lastly, the case is made for SR methods to avoid becoming fractured through being split into different 'types' of review that endeavour to answer different 'types' of question (about effectiveness, or implementation). It is argued that the development of SR methods should be fully cognizant of the contribution of a realist approach that stresses epistemological breadth and ontological depth.

1.0 Critical reflections on research methods

The emphasis of the reflections presented here will primarily be with regard to the use of participant-observation as the key method utilized in this research. In these reflections, I consider the effect that some of the tensions which arose through the use of participant-observation and interviews had upon this research. This enable me to make some suggestions, with the benefit of hindsight, as to how these tensions could be addressed in future research. Given the nature of the funding of this research (a CASE studentship), I also take the opportunity to reflect briefly upon how this mode of funding shaped the research presented in this thesis.

As might well be expected in research that utilizes participant-observation, a tension existed between my insider and outsider roles when present in a research capacity at the institution. Despite my nominal position being purely that of 'researcher', I often felt that my role was quite ambiguous and that there was no clear delineation between my doing work 'for' the institution (such as contributing to meetings, or appraising papers for review) and doing work 'upon' the institution (such as discussing review methods or finding out about institutional procedures) (RD 897-904 15/05/06). My decision to 'strategically' remove myself at times from the immediate environs of the managing team (MT) (by working at a 'hot desk' at the other end of the office) was rooted in my concerns about the team becoming tired of my presence and being unsure about the exact role I was playing (RD 889-904 15/05/06; 5365-5388 15/05/06). However, I also became aware that the MT were conducting their own research upon me; how was I coping with the critical appraisal and data extraction in the course of acting as a co-reviewer? Was I sufficiently acquainted with the scope and methods manual to conduct the work to a satisfactory standard? (RD 5390-5400 17/05/06). This eased my concerns somewhat as it meant that in many ways I was being treated the same as any other team member, although as I noted at the time, I sometimes felt happier wrestling with the review process away from the view of others (RD 5401-5404 17/05/06).

My role as a participant-observer required that I balance the privileged access that I had to the institution with maintaining a constructively critical approach in the conduct of the research. In common with any ethnographic study, this required that I was able to act according to the *mores* of the institution whilst maintaining the ability to step back in order to consider phenomena from a critical perspective outside of the institution. At times, such as when unguarded remarks were made in meetings, this required simply that I acknowledge that I knew I had privileged access; my role was not to highlight controversy but to present a more rounded analysis of the review process (RD 4499-4561 24/03/06). At other times, such as when reflecting upon the process of

critically appraising each paper for the substance misuse review (RD 781-844 15/05/06, *passim*) maintaining a critical approach proved very tiring, as it effectively required that I appraise research papers according to one model whilst simultaneously reflecting critically upon that model in order to inform the case study research (RD 1124-1140 18/05/06).

The issue of privileged access to the institution meant that, despite the efforts made in designing the research in such a way as to satisfy the NHS LREC, ethical 'grey areas' remained. These issues relate to the nature of participant-observation, where conversations may be overheard that substantively inform the research but for which specific consent to use has not been granted (such as would occur in, for example, an interview). In the event, no such events took place in the conduct of this case study. However, I found the lack of clarity over how to maintain ethical conduct should the situation have arisen troubling, especially in the context of the detailed documentation that I had prepared to inform participants in the research (RD 5049-5062 09/05/06). In view of the importance attached by the NHS LREC to participants' right to withdraw from a study, I also felt troubled by the difficulty that a team member would have in withdrawing their consent to take part in the ethnographic aspect of the research[96] (RD 5064-5080 09/05/06). Weighed against this concern was the social position and experience of the people being studied; the team members were themselves members of a research community and their employment within the institution necessarily involved precision in how they presented themselves and their views (RD 5102-5114 09/05/06). Nevertheless, this does not exempt the researcher from endeavouring to adhere to ethical research principles at all times.

[96] Whilst my NHS LREC application clearly defined the procedure for interviewees to withdraw their consent to participate in the study (or to grant consent only for specific sections of their interview to be used), I felt unable to present the participant-observation aspect of the research in such a way as to satisfy the LREC's ethical criteria. This aspect of the research, in terms of explicit ethical approval, was therefore 'glided over' on the basis that ethical practice can be better achieved by adhering to principles of practice rather than explicit procedures (British Sociological Association, 2002).

Despite the problems identified above with regard to conducting participant-observation, it needs to be acknowledged that the method involved me in the work and daily life of the institution in a way that would not have been possible using any other method. This meant that the process of interviewing team members, whilst complicated in some ways by the ambiguous nature of my role in conducting participant-observation[97], was also facilitated by the fact that I was interviewing people that I had gained a good deal of respect for in the course of participating in the work of the SR. Another dimension to this was the empathy that I felt with interviewees regarding the competing demands upon one's time when conducting a review, and the knowledge that it is highly unlikely, if not impossible, to conduct a review that will satisfy all interested parties. The degree of identification that I felt with interviewees should not be viewed as inhibiting critical discussion within the interviews; rather, the mutual respect between interviewer and interviewee may be argued to have facilitated rather than inhibited this critical discussion. In short, more was gained than lost through the use of participant-observation[98].

It is also appropriate at this juncture to consider the role played by the mode of funding for this research (a CASE studentship). The institutional partner, in this instance, played a key role in shaping the direction of the research. This shaping occurs in both a practical sense (for example, participant-observation would not have been possible had the institutional partner not granted access) and in the sense of framing the focus of the research. For example, I have endeavoured in this research to conduct a critical analysis of the process of conducting a SR that draws upon both methodological critiques and discourse analysis; without the active involvement of the

[97] This ambiguity extended to how I viewed my role in relation to the institution itself; as time progressed, I became more aware of how the institution might view me in the future with regard to the utilization of my research or the potential for employment. In this way, I was critiquing the institution's practices (which necessarily involved challenging the power structure) whilst at the same time becoming involved with the power structure myself. (With thanks to Dr Sara Shaw (University College London) for the opportunity to discuss and reflect upon this issue).

[98] This does not mean that participant-observation is essential to conducting research into the process of conducting a SR. However, it does indicate that a greater familiarity with the institutional context in which a review is conducted, in addition to an appreciation of the day-to-day realities of organizing and conducting a review, can play a very important role in contributing to a balanced understanding of the review process.

institutional partner in organizing access to the research area and encouraging my engagement with the development of SR methods, the research may well have focused on just one of these two aspects. I am firmly of the view that the research has greatly benefited from this dual-focus, and moreover has provided an opportunity to make judicious use of discourse analysis in a context directly relevant to policy and practice. This contrasts markedly with my concerns earlier on in the research process, where I struggled to foster the development of the research in a way that was satisfactory for both institutional and academic partners (RD 4593-4613 31/03/06; 5302-5329 15/05/06; 5133-5154 31/05/06).

2.0 Research validity as a framework for conducting systematic reviews for social policy

The difficulties encountered in applying the hierarchy of evidence to a SR in the field of substance misuse has been a recurring theme within the analysis conducted for this research. I have argued that respondents encountered persistent problems in attempting to describe and justify their proposals for methodological development in the terms of the dominant discourse about SR methods. This section draws upon the work of Donald Campbell and colleagues to propose a wider conceptualization of research validity that may be utilized as a framework for conducting SRs for social policy. The framework is not tied to a particular SR method; rather, it is a means of positing questions about what is required of a review and how the review might best proceed to provide rigorous answers to these questions.

Designing a SR around the framework provided by Campbell's three forms of research validity (internal, external, and construct) requires that greater consideration be given to the form of validity that should be emphasized given the nature of the review question, the field of study (can experimental conditions realistically be established?), and the level of development of knowledge in the field (does a consensus exist regarding how phenomena are classified and how mechanisms

operate between them?). The answers to these questions should inform the focus of the SR, both in terms of the subject matter and the criteria by which research will be judged for its rigour. These considerations would substantively guide the emphasis given to the different forms of research validity. Simplified, three situations could arise at the outset of a SR:

1) The results of evaluations in the field of study are equivocal, but there is broad agreement regarding the causative mechanisms at play and the means by which confounding factors can be reasonably controlled. A meta-analysis can be conducted in order to cumulate this knowledge and make the best estimate of effect sizes (internal validity to be prioritized in the SR).

2) There is consensus in the field of study that intervention x is effective in (for example) the United States of America (US). It is not necessary to demonstrate this effectiveness with a greater degree of accuracy, but it *is* important to establish if the intervention would be effective in the United Kingdom (UK) (external validity to be prioritized in the SR).

3) There is contention in the field regarding the manner in which phenomena are conceptualized. Conducting a SR based upon these contested issues risks producing knowledge that is founded upon inadequate constructs (construct validity to be prioritized in the SR).

The rationale advanced above should not be taken to mean that a SR can legitimately focus upon just one type of research validity. Discriminating between the different forms of validity is a useful *tool* for better understanding how to design and conduct a SR, but doing so should not distract attention away from the necessity of understanding research validity as a whole. In Chapter 2 (section 3.1), the manner in which there is a tension between the different forms of validity was discussed with

regard to attaining the best possible balance between them. This conceptualization is certainly a useful tool for visualizing the way in which it is necessary, in any research study or SR, to make trade-offs between the forms of validity in order to focus upon the primary research or SR question. However, it is important to consider whether or not this conceptualization is wholly adequate for understanding how best to balance the forms of validity in a SR for social policy.

To understand how forms of research validity might best be balanced, and moreover to gain an appreciation of *why* internal validity has been so strongly emphasized in Evidence-Based Policy and Practice (EBP) discourses and the traditional SR approach, it is necessary to develop the model of how the forms of validity inter-relate. The development of this understanding is crucial in order to appreciate the way in which ontology underlies all claims to knowledge, and subsequently for gaining an appreciation of other's epistemological positions. This develops points that were made regarding idealist and realist ontologies in Chapter 1 (section 1.5). For the idealist, external validity is 'nested' in internal validity; if results from studies with high internal validity are consistent, then it is considered more than likely that these results will also have external validity (Egger et al., 2001a; Glasziou et al., 2001). Stated another way, internal validity is considered to be fundamental to external validity, whereas the converse is not the case. It is arguably for this reason that such importance has come to be placed upon internal validity in discourses surrounding the rigorous conduct of SRs.

The realist position on validity, which emphasizes the identification of mechanisms and emergent properties of phenomena, allows for a greater parity between internal and external validity. The assessment of emphasis is to be decided upon a case-by-case basis. From the position of the realist, external validity *can* be (although is not necessarily) fundamental to internal validity; for example, 'middle range

theories' about mechanisms identified in one field of study may be judiciously applied to another field in order to inform the manner in which internal validity is assessed.

Very importantly for both the idealist and the realist, it may be argued that *construct* validity is the most fundamental form of validity. If constructs are inadequate, then the knowledge produced in the research study or SR is built upon insecure foundations. No matter how strong the internal validity of a study, if it is based upon inadequately developed constructs its results will be only partially valid[99]. Arguably, it is of prime importance that discourse about SR methods develops so as to acknowledge the fundamental nature of constructs in attaining research validity. This would facilitate the development of a process for debating the validity of constructs that underlie SRs so that these constructs are explicitly acknowledged, rather than relying upon assumptions about constructs.

If the above analysis is correct, at least in part, then it would point towards the need to develop models for conducting SRs for EBP that have both epistemological breadth and ontological depth (Bhaskar, 1978; Sayer, 1992; 2000, p.10-28; Byrne, 2002, p.12-28 and 79-94; Clegg, 2005; Pawson, 2006, p.17-37). The epistemological breadth in such a SR would come from a judicious, case-by-case approach to how different forms of knowledge can contribute to the evidence-base; and the ontological depth from the ability to provide an adequate account of why what has been observed to happen at certain times is taken to apply to future outcomes regarding the phenomena concerned. As noted above, research validity needs to be considered as a whole. The distinctions between internal, external, and construct validity can begin to feel somewhat forced if they are treated in isolation, but their usefulness as a framework to structure thinking about and discussing the issues involved in a piece of research or a SR arguably remain.

[99] This would not come as any surprise to statisticians, nor indeed to meta-analysts, who are only too aware of the dangers of 'garbage in garbage out' (Cooper and Hedges, 1994). However, the question remains as to how willing researchers (as well as systematic reviewers and policy makers) are to critically reflect upon the basis of the constructs which they utilize.

Understanding research validity as a whole, rather than compartmentalized areas, has significant implications for the way in which different types of questions are to be 'answered' by SRs. It is erroneous to think that we can 'split off' evaluations of effectiveness from knowledge about the implementation of interventions. It is not the case, in any area of social policy, that a SR prioritizing internal validity can simply be complemented by another SR prioritizing external validity *unless* a substantive ontological model can be provided to account for this division of knowledge. Unfortunately, the development of SR methods in certain areas is proceeding upon the basis of these shaky foundations (for example, see Brunton et al., 2005; Shepherd et al., 2006; Arai et al., 2007; Pope et al., 2007, p.9-15). The three forms of validity are present in *every* piece of research and therefore need to be considered as a whole; for example, constructs substantively inform the manner in which concepts are operationalized in a study with high internal validity, and in a similar manner, research prioritizing the testing of different constructs about phenomena cannot afford to overlook paying substantial attention to internal validity. A SR can only have rigour in the syntheses that it provides on the basis of attaining a certain degree of validity in *all three* areas of research validity, the emphasis upon each being agreed upon at the outset, or renegotiated in the light of findings made in the course of conducting the review.

In the final analysis, in the case study in this research, the implementation of a particular SR model can be seen to have not straightforwardly directed an 'evidence-based' policy making process. In part, this is arguably because the policy makers concerned demonstrated the same critical approach to the knowledge produced in the substance misuse review as they would to knowledge produced by other sources; although proficient at engaging in EBP discourses, the policy makers did so in a measured way that demonstrated an ongoing critique of the methods used and a desire to foster the development of methods so as to better inform the committee's deliberations. Nevertheless, the significant role played by both EBP and substance

misuse discourses in directing and constraining the boundaries of what it was possible for policy makers to think, whether with regard to SR methods or the way in which the review and policy were framed, should not be underestimated. However, it can be confidently stated that the problems encountered in implementing the SR methods in this case study have played a part in stimulating the development of a range of research synthesis methods that will themselves require testing out in order to evaluate their strength and weaknesses in producing knowledge that can inform policy and practice.

A commitment to the ongoing development of SR methods for social policy is plainly desirable. The case study in this research has highlighted the problems that exist regarding the volume of evidence that a SR can produce and the substantial difficulties that can result in endeavouring to synthesize and deliberate upon it. There exists the distinct danger here that SRs will fail to inform policy making any better than the (unformulated) methods utilized prior to the institutionalization of EBP. There is even the risk that policy could become *less* informed by evidence; if a SR produces knowledge that is indistinct and which is highly problematic to interpret, policy makers are obliged to fall back onto common sense or ideological conceptualizations of, and solutions to, complex social problems. In the final analysis, we are still reliant upon expertise to interpret knowledge for policy and practice. Making explicit the conceptualizations and underlying knowledge that experts utilize, whether writing an 'academic' or narrative review or deliberating upon a SR in order to make policy recommendations, is more important than attempting to remove the role played by this expertise in a bid to attain 'objectivity'. Reflecting upon the way in which discourses about SR methods and the topics they investigate affect the way that these methods and topics are conceptualized, positioned, and discussed, could substantively inform the methodological development of these methods and thereby contribute to producing knowledge that better informs policy and practice.

Appendix A

Research Protocol:
The Construction and Utilization of an Evidence-Base

Chief investigator:
Mark Pearson, School of Sociology, Politics and Law, University of Plymouth, Drake Circus, Plymouth. PL4 8AA.
E-Mail: mark.pearson@plymouth.ac.uk
Office: (01752) 238567 Mobile: (07876) 687433 Home: (020) 7687 6201

Research supervisors:
Dr Ross Coomber, School of Sociology, Politics and Law, University of Plymouth, Drake Circus, Plymouth. PL4 8AA.
E-Mail: ross.coomber@plymouth.ac.uk
Office: (01752) 233218

Professor Mike Kelly, National Institute for Health and Clinical Excellence, MidCity Place, 71 High Holborn, London. WC1V 6NA.
E-Mail: Mike.Kelly@nice.org.uk
Office: (020) 7061 3150

This research is funded by a Collaborative Awards for Science and Engineering (CASE) studentship agreed between the Economic and Social Research Council, the Health Development Agency (whose functions are now taken on by the National Institute for Health and Clinical Excellence) and the University of Plymouth.

Rationale and Background Information

The growth of demand for evidence-based policy, and in turn for the evidence-bases upon which policy decisions may be made, has been significant in the UK in recent years. Whilst most past public policy would no doubt lay claim to being based, at least to some extent, on 'evidence' and a rational policy approach that took the available evidence into consideration, current calls for evidence-based policy are in part a result of the failings of previous public policy and claims from critics of the relatively uninformed and partisan nature of much policy formation. One contemporary manifestation of this demand for evidence of 'what works' is the Centre for Public Health Excellence (CPHE) at the National Institute for Health and Clinical Excellence (NICE). The remit of the CPHE is to identify the evidence of what works to improve people's health and reduce health inequalities.

Explicit in the call for evidence-based policy is the idea that objective evidence that is based on what has been shown to be effective is essentially about systematically and appropriately collating, weighting, assessing and analysing the available data and coming up with either proven (or demonstrable) 'best practice' or the most rationale approach to the issue in hand – as directed by the evidence.

Study Goals and Objectives

The mechanics of finding, reviewing and synthesizing research evidence have been extensively investigated in the 'evidence-base' literature. However, contention persists over the best methods for integrating diverse forms of knowledge for application in professional practice and policy-making. In a complex field, such as the prevention of substance misuse, this integration of knowledge is crucial for the pragmatic application of the best evidence.

This research will investigate the manner in which an evidence-base on substance misuse prevention was constructed, namely the evidence-base on community interventions to reduce substance misuse among the most vulnerable and disadvantaged young people (hereafter, 'the evidence-base').
It will seek to make clearer the 'hidden' aspects of the construction of this evidence-base; this will involve not just the investigation of the methods by which the evidence-base was constructed, but also the initial nature of the commissioning of the review and how this affected the outcome of the final evidence-base. In this manner, a broader understanding of the production of the evidence-base would not simply be a matter of assessing the ways that the producers are approaching the technicalities of method. Rather, it would also contribute to an understanding of why particular evidence was sought in the first place and what this means for an understanding of the issue in hand and those that the newly constructed evidence base would affect. In this way, the research will be able to contribute to substantive discussion upon how policy or other social conditions may drive not just a substance misuse prevention evidence-base, but evidence-bases more generally. It will also facilitate patient and public involvement through making the more 'hidden' aspects of evidence-base production explicit and thus open for discussion.

Research Design and Methods

The chief investigator will initially be based at the CPHE at NICE in order to gain familiarity with the day-to-day working life of the Centre; this will be obtained through attending meetings and shadowing key members of staff. The chief investigator will be present as a participant-observer at the key stages of the evidence-review process:

- Review of stakeholder comments and revision of evidence-base scope
- The evidence-review and evidence-synthesis process
- PHIAC meeting to discuss the draft of the evidence-base
- Fieldwork meetings at which the evidence-base will be 'tested' with stakeholders
- Revision of guidance on basis of feedback from fieldwork meetings
- Publication of the evidence-base

Structured field notes will be kept regarding these stages in order that they can be reflected upon for the purposes of topic development for the subsequent semi-structured interviews.

Access to the Centre will be negotiated with, and supervised by, Professor Mike Kelly (Director of the Centre and the institutional supervisor for the Research), with line management the responsibility of Simon Ellis (Associate Director, Methodology). Stages of the evidence-review process will be undertaken by the National Collaborating Centre for Drug Prevention (NCCDP, based at Liverpool John Moores University); access in this instance will be negotiated with Professor Mark Bellis (Director of the Centre for Public Health). The nature of the Research research will be fully explained to all staff members whom the chief investigator interacts with in order that observation is not undertaken covertly. Research staff at both Centres involved in the process of constructing the evidence-base will be recruited for in-depth, semi-structured qualitative interview. This research design allows for a flexible approach in which the chief investigator's time at the Centres can usefully feed into the refinement of the interview design, whilst also allowing for important emergent issues to be incorporated into the research framework.

With regard to researching the wider issues surrounding the commissioning of the evidence-base, contact with non-NHS civil servants involved in the commissioning of the evidence-base, will also be developed. These working relationships will be facilitated and supervised by Professor Mike Kelly, Director of the CPHE (NICE).

The interviews conducted will be semi-structured in nature. It is envisaged that the interview topic guide will be developed upon the basis of reflection on the chief investigator's structured field notes and the substantive content of other interviews. The aim of the interviews is to retain a focus upon certain issues (as outlined in the interview topic guide) whilst also retaining sufficient flexibility to allow interviewees to explore these issues and introduce novel topics if these are relevant. The chief investigator, in the role of interviewer, will place emphasis upon cultivating a social situation whereby participants feel confident in exploring the issues; the interviewer will ensure that this is attained by paying close attention to what the participant is saying and taking a considered approach to probing the issues raised.

Duration of Project

The chief investigator's familiarisation with the Centre and the initial stages of structured observation will take place in April 2006. The chief investigator will act as a participant-observer during the evidence-synthesis process (late April – July), the consultation period (August – September) and the fieldwork period (October – December). Interviews will take place, by negotiation with the research participants, throughout this time period. Data analysis will take place throughout this time and will be completed by the end of March 2007. The thesis will be completed and submitted by December 2007.

Methods of Data Analysis

Data analysis will be organised into three stages:

1) Data management – the non-anonymized interview transcripts will be entered into the qualitative data analysis software, *Nvivo*. Following an initial reading through of the interview, data from the structured field notes and the chief investigator's reflective diary will be attached as 'memos' to sections of the interview data. These 'memos' are intended to add pertinent information that will facilitate later in-depth analysis. At this stage, provisional broad themes within the interview data will also be identified, both as a means of informing later interviews and reflections, and so as to provide a starting point for the subsequent in-depth analysis.

2) Descriptive accounts – the *Nvivo* software will be used as a tool for coding the data (in conjunction with the 'memos' from field notes and the reflective diary) and for presenting this coding in a manner that allows the development of analytical themes and sub-themes within it. The themes and sub-themes will be explored for patterns and associations within and between accounts, in order that further refinement of the themes may be attained. Particular attention at this stage will be paid to describing themes in similar terms to those used by the interviewee, in order that subsequent analysis is strongly grounded in the original data.

3) Explanatory accounts – the patterns and associations of the themes and sub-themes of the data will be analysed in order to develop plausible explanatory accounts of why these patterns and associations occurred. This analytical process will entail a repeated going back and forth between the themes/ sub-themes and the data in its original context, so as to allow the validity of the analysis to be checked and to facilitate further refinement of the analysis. At this stage, the analysis will be required to account for *all* relevant instances if it is to be considered as valid ('comprehensive data treatment'). The existence of discrepant cases, however, should not necessarily be viewed as invalidating the analysis totally; there is considerable scope for further analysis of these cases for the explanatory insights and contribution to theory development which they may provide ('deviant case analysis').

The process of interviewing, transcription and analysis is not a purely sequential. For example, the initial organization of data from one interview may cast new light onto data from a previous interview which is at a later stage of analysis, requiring it to be re-analysed. This going backwards and forwards through the data increases the validity of the research by endeavouring to provide a comprehensive explanatory account of all of the data. The use of *Nvivo* significantly aids this process through providing a quick and reliable tool by which the data may be recalled for analysis according to different hypotheses; for example, thematic categories across different cases can be investigated, or an alternative analysis of the data within one case can be performed without losing the original analysis.

The rigour of the data analysis will be ensured in two ways. First, the initial analysis of the data will substantively inform the subsequent interviews that take place. Secondly, the supervisory team will discuss samples of the data analysis with the chief investigator, and will provide a supervisory framework which ensures the rigour of the data analysis.

Access to Research Findings

The findings of the research will be submitted as a poster to be considered for presentation at the NICE Conference (December 2007). The final report will be available to the research participants in the following forms:
a) the thesis itself; as per UK University procedures, a copy will be lodged in the British Library
b) a final report to be submitted to the Centre (as per the Intellectual Property Agreement dated 29/10/04); it is also envisaged that a presentation will be made at the Centre
c) publication in peer-reviewed journals (*Evidence & Policy; Addiction; Sociology)*

Ethical Issues

All potential participants will be provided with an information sheet providing a clear overview of research topic and the proposed manner in which they could be involved. The voluntary nature of the participants' involvement will be highlighted, along with the sources of funding for the research. The potential benefits of taking part in the research, through advancing knowledge relating to evidence-bases, will also be provided in the information sheet. The information sheet also contains a section regarding the risks of the research regarding data management, and the steps taken in the research protocol to minimize these. The building of open and trusting working relationships with potential participants is a key part of this process of information-giving, and is intended to ensure that the giving of informed consent is not confined to obtaining a signature.

Participants will have the right to withdraw their consent to participate in the research at any time. Participants will also be provided with the contact details of the chief investigator's supervisor, in the event that they have concerns over the conduct of the research. It will also be made clear to participants that the research is conducted within the guidelines of the NHS Research Governance framework and the University of Plymouth's Code of Good Practice in Research. The University's Code also contains an explicit protocol for the investigation of research misconduct.

It is envisaged that interviews of NICE staff will take place in a private room at the Centre's offices in London. The provision of space for this purpose is assured by the terms of the CASE studentship under which this research is funded. It is recognized that some members of staff may spend extended periods away from the London offices; in this instance, and in the case of all non-NHS staff, the chief investigator will travel to the location at which it is most convenient for the person to be interviewed. In

this instance, a private space in which the interview can take place will be organised by mutual agreement between the chief investigator and the interviewee. Interviews will be arranged to take place at a time of mutual convenience to the chief investigator and interviewee; a digital recorder will be used to record the interview, with a back-up on analogue tape also being taken. The analogue copy will be erased once a successful, full transcription of the electronic recording has been made.

It is acknowledged that ensuring absolute anonymity in this research may be problematic; the small number of people who were involved in the production of the drug misuse prevention evidence-base may mean that it is possible to infer the identities of participants from the views that they express (and which are reported in the final write-up). The following steps will be taken in an effort to assure anonymity:

1) Participants will have the right to withdraw any statement made in the course of an interview; they will also have the opportunity to do so in the 14 days in which they have to review the transcript of the interview
2) In the conduct of interviews, the chief investigator will not refer to other participants by name if making reference to the discussions that took place in another interview
3) All references to names made in the interviews will be anonymized in the transcripts

In view of the high status of the interviewees, it is considered reasonable to assume that they will be both articulate and capable of organising their time effectively. The potential for interviewees feeling unduly pressurised to participate in the research, or of their participation adversely affecting the work that they are employed to do for the NHS (or other organization or government body), is thus considered to be minimal. Moreover, the topic of the research is one that is likely to be of interest to all those who are involved professionally in the construction of evidence-bases. The research at the Centre is undertaken according to the timetable agreed to in an Intellectual Property Agreement (dated 29/10/04) between the HDA, the University of Plymouth and the chief investigator. This agreement was drawn up in respect of the CASE studentship through which this Research is funded; it should be noted that the responsibilities of the HDA passed to NICE when the organizations merged in April 2005.

The management of data in the course of this research will be explicitly addressed on the participant consent form; in particular, the details of the security of the data storage are clearly provided and permission sought for the archiving of the data through an institution which has transparent policies for the responsible secondary analysis of data (the *Economic and Social Data Service* (University of Essex)).

Explicit permission will be sought on the consent form for storing the participant's name and position together with the (non-anonymized) transcript for the duration of the data analysis by the chief investigator. This is to facilitate an informed analysis of the data, where the chief investigator's knowledge of exactly who is being referred to may be considered vital. An option will also be offered on the consent form for participants to give their permission for anonymity not to be preserved; it is important to remember that the topics to be discussed are not necessarily ones that participants will want to remain hidden. Participants may find the interview a constructive process that contributes to learning and best practice within the NHS, and see no need for their identity to be hidden.

Appendix B

Information Sheet: 'The Construction and Utilization of an Evidence-Base'

This study forms part of the doctoral research to be conducted by Mark Pearson (University of Plymouth). This information sheet is designed to provide the information you will require to make an informed decision as to whether or not you wish to participate in the research.

The Purpose of the Research

This research forms part of a collaboratively funded studentship that was originally negotiated between the Economic & Social Research Council, the University of Plymouth, and the Health Development Agency (HDA). Following the integration of the HDA into the National Institute for Health and Clinical Excellence (NICE), the studentship will be completed under the auspices of NICE. The research will utilize the evidence-base on community interventions to reduce substance misuse among the most vulnerable and disadvantaged young people (hereafter, 'the evidence-base') as a case study to investigate the manner in which evidence-bases are constructed and utilized. It will seek to make clearer the 'hidden' aspects of the construction of the evidence-base; this will involve not just the investigation of the methods by which the evidence-base was constructed, but also the initial nature of the commissioning of the review and how this affected the outcome of the final evidence-base.

Participation in the Research

People who have been involved with the commissioning and conduct of the evidence-base will be purposively selected to take part in this research. The earlier stages of the research, where the researcher is acting as a participant-observer in meetings and in the day-to-day work of the review, will involve the researcher recording his observations in the form of a research diary. These observations will be used to structure reflection upon the review process in order to develop topics for discussion at in-depth interviews with key members of the team conducting the review.

The research diary may be viewed at any time (as a Word document) by members of the team conducting the review. It is stressed that the research diary will not be used in order to provide a direct record of events or quotations for the purposes of the thesis; its role is to facilitate the critical development of ideas on the part of the researcher. At the interview stage, participants will be asked to sign a consent form in which consent can be given (or withheld) for the use of direct quotations in the thesis. These interviews will be digitally recorded (together with a back-up recording on analogue tape) and transcribed by the researcher; the transcription will be returned to the interviewee for checking and to allow for the option to withdraw statements to be taken up if desired.

It is intended that the experience of participation in this research should be a constructive one that contributes to learning and best practice within the NHS. However, it is acknowledged that contentious issues may be discussed in the course of the interviews; the following section provides details regarding the guarantee of confidentiality on the part of the researcher, the preservation of anonymity, and data management procedures. This research is governed by the University of Plymouth's Code of Good Practice in Research (2002), which also contains an explicit protocol for the investigation of research misconduct. You are encouraged to contact and discuss with the researcher any concerns that you may have regarding the conduct of this research.

Research Data Management

The researcher, Mark Pearson, is bound by the University of Plymouth's Code of Good Practice in Research to maintain confidentiality regarding all issues discussed in the course of this research, *unless* the participant explicitly states that they are satisfied to be identified with a particular statement.

It is acknowledged that ensuring absolute anonymity in this research may be problematic; the small number of people who are involved in the production of the evidence-base may mean that it is possible to infer the identities of participants from the views that they express. If you are happy for your identity to be known in the report of the research, there is an option on the consent form for you to grant permission for this.

All interviews undertaken in this research will be transcribed within 5 days of the interview taking place and e-mailed to you so that the transcription may be checked. It is requested that you

complete this checking within 14 days of the transcription being sent; the opportunity exists at this stage for you to withdraw, or clarify, any statements you have made.

Data Storage

The digital recordings of interviews, and the verbatim transcriptions thereof, will be stored electronically on the researcher's laptop computer, on the University of Plymouth's intranet server, and as a back-up on a USB memory stick and/or CD. The analogue recordings will be erased as soon as a satisfactory transcription has been made from the digital recording. The recording and transcription files will be password-protected using the Microsoft XP operating system. In view of the importance of the contextual information required by the researcher for data analysis, the interviews will NOT be anonymized at this stage. However, anonymisation will be performed for the purposes of any quotations utilized in the report of the research.

In accordance with the University of Plymouth's Code of Good Practice in Research, in order to maintain the transparency of the research the password-protected computer files of the *anonymized* transcripts and digital recordings will be stored for a minimum of 10 years. The contents of the researcher's fieldwork diary (providing contextual details of, and reflections upon, the interview) will also be stored for this time period. The files containing the non-anonymized transcripts will be deleted upon completion of the Ph.D. The anonymized transcripts and the contents of the researcher's fieldwork diary will be submitted to the *Economic and Social Data Service* (University of Essex) for consideration for archiving; if the data are accepted, their responsible use will be governed by an end user licence.

Access to, and Use of, the Data

Access to the non-anonymized data will be restricted solely to the researcher, Mark Pearson. The anonymized data may be shared, in the pursuit of normal discussions regarding doctoral research, with the researcher's Ph.D. supervisors: Dr Ross Coomber and Professor Malcolm Williams (both University of Plymouth), and Professor Mike Kelly (NICE). Upon completion of the Ph.D., the anonymized transcripts may be deposited with the *Economic and Social Data Service* (University of Essex) so that other social researchers may access them to perform re-analysis or verification.

In addition to being written up in the form of a Ph.D., it is also proposed that the data analysis be utilized in writing papers for publication in peer-reviewed journals (*Evidence & Policy; Addiction; Sociology)*. The analysis will also form the basis of a report to be submitted to NICE as part of the agreement of funding the studentship. The findings of the research will also be submitted as a poster to be considered for presentation at the NICE Conference in December 2007. As would normally be expected from Ph.D. research, the data analysis may substantively inform an application by the researcher for a Post-Doctoral Fellowship, or for applied research within the NHS.

Ethical Approval

This study was given a favourable ethical opinion for conduct in the NHS by the UCLH Research Ethics Committee.

If you have any further questions, please contact either the researcher or a member of the supervisory team:

Research Student - Mark Pearson, School of Sociology, Politics and Law, University of Plymouth, Drake Circus, Plymouth, Devon. PL4 8AA.
E-Mail: mark.pearson@plymouth.ac.uk Tel: (01752) 238567 or NICE: (020) 7400 0669

Supervisory Team
Dr Ross Coomber, University of Plymouth
E-Mail: ross.coomber@plymouth.ac.uk Tel: (01752) 233218

Professor Mike Kelly, National Institute for Health and Clinical Excellence
E-Mail: Mike.Kelly@nice.org.uk Tel: (020) 7061 3150

Appendix C

Research Protocol for PHIAC Meetings: The Construction and Utilization of an Evidence-Base

Ph.D. student:
Mark Pearson, School of Sociology, Politics and Law, University of Plymouth
E-Mail: mark.pearson@plymouth.ac.uk Mobile: (07876) 687433

Research supervisors:
Dr Ross Coomber, University of Plymouth
E-Mail: ross.coomber@plymouth.ac.uk Office: (01752) 233218

Professor Mike Kelly, National Institute for Health and Clinical Excellence
E-Mail: Mike.Kelly@nice.org.uk Office: (020) 7061 3150

This research investigates the process by which an evidence-base (community interventions to reduce substance misuse among the most vulnerable and disadvantaged young people) is produced by NICE. The investigation of this process necessarily includes research upon the process of searching for, screening, and appraising papers (and extracting data from them). However, this research also investigates the processes involved in deriving evidence statements from this data, and the manner in which these statements are utilized (at the PHIAC and fieldwork meetings) in order to produce guidelines for policy and practice.

The research has involved participant-observation in all stages of the review to date; the initial stakeholder meeting, revision of the scope, the screening and appraisal of papers and the extraction of data for the review, and the meetings and teleconferences in which the planning and monitoring of the review took place. The research protocol for the whole study is available on request from Mark Pearson.

The objective of observing the PHIAC meetings is to examine the way in which evidence is utilized in the process of argumentation, and how this builds into the production of (draft) guidelines. Bales' method of Interaction Process Analysis will be used as the basis for classifying who speaks during the meeting, about what, to whom, and with what response. The aim is not so much to simply categorize these processes, but to provide the basis for a contextualized analysis of the manner in which argumentation (point, counter-point, and consensus-building or disagreement) proceeds.

The anonymity of all participants in the PHIAC meeting will be maintained in the course of conducting this research. The matters discussed will be treated in confidence, and whilst notes on the details of the points discussed during the meeting shall be made, no direct quotations will be recorded. Dissemination of this research will be in the form of a Ph.D. thesis and a report submitted to the Centre for Public Health Excellence at NICE. The research is conducted within the NHS Research Governance framework and abides by the University of Plymouth's Code of Good Practice in Research.

Appendix D

Interview Topic Guide
(for team members involved in substance misuse review)

1. The scope and stakeholder consultation
   - setting the initial review question
   - the experience of consulting stakeholders
   - revising the scope
   - using the scope to conduct the review

2. The review itself – strengths and weaknesses, successes and problems
   - searching
   - screening
   - appraising
   - extracting data
   - synthesizing/ writing-up

3. Managing the review process
   - principles and key considerations
   - timelines
   - in comparison to other reviews
   - changes resulting from new institutional arrangements

4. Committee meetings and the drafting of recommendations
   - utilization of the effectiveness and cost-effectiveness reviews
   - utilization of the additional information and analyses presented at the meeting
   - utilization of co-optee expertise

5. Any other topics that interviewee would like to address

Appendix E

Interview Topic Guide
(for team member with responsibility for economic modelling)

1. The cost-effectiveness review itself – strengths and weaknesses, successes and problems

2. The cost-effectiveness model
   - Assumptions
   - Dealing with complexity
   - Generalising to the UK context
   - Intervention effectiveness and cost-effectiveness – whether to consider separately or in an integrated manner

3. Managing the review process
   - principles and key considerations
   - timelines
   - in comparison to other reviews

4. Committee meetings and the drafting of recommendations
   - utilization of the cost-effectiveness review
   - utilization of the additional information and analyses presented at the meeting
   - utilization of co-optee expertise

5. Any other topics that interviewee would like to address

Appendix F

Interview Topic Guide
(for senior team member)

1. The scope and stakeholder consultation
   - setting the initial review question
   - the experience of consulting stakeholders
   - revising the scope
   - using the scope to conduct the review

2. The review itself – strengths and weaknesses, successes and problems
   - screening
   - appraising
   - extracting data
   - synthesizing/ writing-up

3. Committee meetings and the drafting of recommendations
   - utilization of the effectiveness and cost-effectiveness reviews
   - utilization of co-optee expertise
   - working with the limitations of the evidence-base

4. Changes resulting from new institutional arrangements
   - changes to the review process
   - timelines
   - methodology - the 'review itself' and processes surrounding the review
   - the integration of cost-effectiveness

   - changes to the:
   - management of the review process
   - roles played by staff
   - field of Public Health

   - current methodological development of review process

5. Any other topics that interviewee would like to address

Appendix G
Consent Form

ID number:

Project title: 'The Construction and Utilization of an Evidence-Base'
Researcher: Mark Pearson
Institution: School of Sociology, Politics and Law, University of Plymouth, Drake Circus, Plymouth, Devon. PL4 8AA.
Tel:(01752) 238567 (office), (07876) 687433 (mobile) **E-Mail:** mark.pearson@plymouth.ac.uk

1. I confirm that I have read, and understand, the information sheet (v.1.1) for the above study. I have had the opportunity to consider the information and ask questions, and have had these answered satisfactorily.
2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason.
3. I understand that both a digital and analogue recording will be made of any interviews which I hold with the researcher, and that i) the analogue recording will be erased immediately a satisfactory transcript has been made, and ii) the digital recording will be kept for a period of at least 10 years (with the file password-protected and no identifying data being held in the same location) in accordance with'the University of Plymouth's Code of Good Practice in Research (2002).
4. I understand that a transcript of my interview will be sent by e-mail to me within 7 days of the date of the interview. I then have a period of 14 days in which to comment upon, clarify, or withdraw statements from the transcript. I also understand that, upon request, I can be sent by e-mail the digital recording of the interview.
5. I understand that the researcher will retain a *non*-anonymized transcript of my interview for the duration of the data analysis for the purposes of the Ph.D. This file will be password-protected, and all copies of it deleted upon completion of the Ph.D.
6. I understand that the information I supply will normally be preserved by the *Economic and Social Data Service* (ESDS) and will be kept confidential unless I give permission for my name to be used. Offering this data to ESDS is encouraged by the ESRC in order to make the data publicly accessible for critical appraisal and re-utilization. The contents of the researcher's fieldwork diary (providing contextual details of, and reflections upon, the interview) will also be offered to ESDS for archiving. The material will be preserved as a permanent research resource for use in research and publication, subject to the terms and conditions of an end-user licence (see: http://www.esds.ac.uk/aandp/access/summary.asp)
7. I understand that the researcher, Mark Pearson, is bound by the Data Protection Act (1998) and the University of Plymouth's Code of Good Practice in Research (2002) to maintain confidentiality regarding all issues discussed in the course of this research, unless I explicitly state that I am satisfied to be identified with a particular statement.
8. I agree to take part in the above study, and hereby assign the copyright in my contribution to Mark Pearson

Please tick either:
□ I give my permission for the information I have given to be used for research purposes only (including research publications and reports) **without** preservation of anonymity
□ I give my permission for the information I have given to be used for research purposes only (including research publications and reports) **with strict** preservation of anonymity
Please tick either:
□ I give my permission for direct quotations from this interview to be made in the write-up of this research
□ I **do not** give my permission for direct quotations from this interview to be made in the write-up of this research

Participant

Name:	Signature:	Date:
.....	.....	.....

Researcher

Name:	Signature:	Date:
.....	.....	.....

Appendix H: Studies Excluded from the Substance Misuse Review: Classified by Reason for Exclusion

For reasons of brevity, excluded studies are cited in this appendix in the form of Author (date of publication) - full citations can be found in the review itself (Jones et al., 2006). The code preceding the citation is the identification number assigned to each study by the review team; it allows the decisions made about each study to be simply located in the review database.

Category (1): Study topic outside of review's scope

37 Knight et al 2005
16698 Anguelov 1999
16278 Gonzalez 1990
15454 Amodeo and Gal 1997
15331 Marlatt et al 1998
15186 Novakova 1999
44 Schinke and Schwinn 2005
5435 Fearnow et al 2003
87 Simons-Morten et al 2005
86 Werch et al 2005
114 Fok et al 2005
15060 Wagenaar et al 2000
14987 D'Amico and Fromme 2000
210 Derzon et al 2005
14875 Andsager et al 2001
14610 Fraguela et al 2003
14300 Salmon et al 2005
487 Orlando et al 2005
14187 Hogue et al 2005
2957 Hansen 1999
593 Grant et al 2004
4471 Andrews et al 1995
4168 Kisker 1996
14173 Flawelling et al 2005
14100 Lalor et al 2006
14072 Low 1990
4204 Hawthorne 1996
4229 Clayton et al 1996
4262 Hahn et al 1996
4599 Nelson-Simley 1995
4656 Hawthorne et al 1995
4724 Anonymous 1994
5034 Elliot et al 2006
408 Faggiano et al 2005
664 Bond et al 2004
5050 Longshore et al 2006
5051 Zand et al 2006
5065 Sale et al 2006
13175 Lister-Sharp et al 1999
13130 Tobler 2000
5066 St Pierre et al 2005
5082 Barnes et al 2004
722 Griffin et al 2004
907 Ghosh-Dadtidar et al 2004
987 Bond et al 2004
1013 McBride et al 2004

1016 Nilsson et al 2004
13028 Noble et al 2000
5044 Slater et al 2006
12909 Starkey and Orme 2001
1051 Furr-Holden et al 2004
1067 Dent et al 2001
1078 Steinman et al 2003
1079 Tait 2003
12806 D'Amico and Fromme 2002
1116 Licciardone 2003
1142 Acierno et al 2003
1145 McBride 2003
1158 DiNoia et al 2003
13247 Walker 1999
13194 Olds 1999
13254 Jones 1999
13287 Atkan 1999
11229 Mohai 1991
11230 Mohai 1991 (duplicate of 11229)
10219 Zagumny and Thompson 1997
10207 Silva et al 1997
10145 Eggert et al 1997
10133 Joyner 1997
1371 Fendrich and Rosenbaum (no date)
1376 Trudeau et al 2003
10027 Rast 1998
1350 Eisen et al 2003
9930 Tobler et al 1999
1386 Yzer et al 2003
1538 Finke et al 2002
1599 Wu et al 2002
1641 Halfors et al 2002
9890 Reuter et al 1999
1677 Cuijpers 2002
1684 Barkauskas et al 2002
15082 Ullman et al 2000
9849 Jackson and Matthews 1999
1916 Cuijpers et al 2002
2045 Carroll et al 2001
2059 Botvin et al 2001
2128 Barrowclough et al 2001
2410 Park et al 2000
2412 Taylor et al 2000
2494 Miller et al 2000
2546 Botvin et al 2000
2614 Lang et al 2000
1744 Eisen et al 2002
2716 Prinz et al 2000
2749 Goldberg et al (no date)
2801 Biglan et al 2000
1793 Kumar et al 2002
1880 Collins et al 2002
10337 Dore 1996
10452 Tashijan 1996
10489 Barker 1995
10696 Romero 1994
10698 Silva 1994

10721 Ertle 1994
10842 Fahs 1993
10850 Brown 1993
10851 Anonymous 1993
10978 Piper 1993
11316 Schinke 1991
11885 Dusenbury 2005
12122 Gerada 2005
12476 Broner 2003
12676 Halfors 2002
12736 Baker 2002
12744 Baker 2002 (duplicate of 12736)
5538 Mathias 1997
5601 Marsh et al 2000
5657 Smitham et al 1999
9847 Catalano et al 1999
6121 Tsemberist et al 2004
6356 Smit et al 2003
6471 Fielding et al 2002
6853 Piper et al 2000
3279 Spoth et al 1999
3307 White et al 1998
7342 Azrin et al 1996
3455 Lisnov et al 1998
2887 Lynam et al 1999
7822 Faggiano et al 2006
9585 Goldberg et al 2001
9838 Harrell et al 1999
3534 Palmer et al 1998
3536 Black et al 1998
31 Borsari and Carey 2005
833 Boekeloo et al 2004
856 Guyll et al 2004
1475 Perry et al 2003
1582 Henry et al 2002
1807 Spoth et al 2002
2386 Beebe et al 2001
3927 Boyer et al 1997
3915 Dusenbury et al 1997
3826 Ashery et al 1997
3821 Biglan et al 1997
7851 Skara 2003
8128 Rohrbach 1993
8753 Murray 1990
8754 Johnson 1990
8755 Botvin 1990
8849 Botvin 1990
9237 Abbey 1990
9349 Spoth 2005
9574 Harrington 2001
9662 Karoly 2001
5185 Riper et al 2005
5379 Ringwalt et al 2003
16777 Berry and McKenna 1993
5210 Pantin et al 2003
15387 Gorman 1996
8517 Coggans et al 1991

16742 Henggeler et al 1996
5782 Branigan and Wellings 1998
5792 Paxton et al 1998
5511 McBride et al 2002
5220 Malcolm et al 2004
6843 Goldberg et al 2000
5223 Haines et al 2005
5238 Pankratz and Hallfors 2004
14659 Pentz and Li 2002
4275 Anonymous 1996
15513 Furlong 1997
5365 Ellickson et al 2003
5337 Montoya et al 2003
5329 Eischens et al 2004
14378 Gould 2004
14778 Fischer-Potts 2002
16537 Irwin 1994
16461 Ferrell 1997
15170 Saunders 1999
15173 Young 1999
16107 Jorgensen 1990
16101 Flisser 1991
15736 Bennett 1995
15721 Baranski1999
14230 Berkley-Patton 2004
547 Russell et al 2005
5685 Johnson et al 2000
5116 Daenseekaew et al 2005
5429 Shaughnessy 2003
1869 Ahmed et al 2002
15955 Baer 1993
15471 Ammerman and Hersen 1997
14491 Dishion and Kavanagh 2003
15710 Bukstein 1995
16167 Berger 1991
2500 Donnermeyer 2000
10749 Bullis et al 1994
14802 Gropper 2002
5029 White et al 2006
2459 Murray et al 2000
11914 Watson et al 2004
1995 Giles et al 2001
12570 Boyd-Ball 2003
14927 August et al 2001
2652 Abbey et al 2000
12720 Belgrave 2002
5466 Holleran et al 2002
16695 Banquer and Hoganbruen 2000
14198 Schinke et al 2005
1083 Sussman et al 2003
1663 Peleg-Oren 2002
5261 Komro et al 2004
5368 Botvin et al 2003
9822 Myers et al 2000
1560 Tarter et al 2002
1556 Gorman-Smith et al 2002
966 Tatchell et al 2004

14276 Javis and Stark 2005
1627 Macaulay et al 2002
946 Paz et al 2004
9791 Risberg and Funk 2000
2648 Byrne et al 2000
2420 Bergamaschi et al 2000
14919 Molgaard and Spoth 2001
1198 Lilja et al 2003
5905 Roberts 1995
5878 Harris and Ludwig 1996
15975 Arella 1993
16499 Ehrensal 1996
13842 Resnicow 1994
16124 Clayton et al 1991
10071 McDaniel and Shreeve 1998
3646 Dukes et al 1997
14261 Elliott et al 2005
9225 Church et al 1990
63 Gray et al 2005
9207 Pentz et al 1990
7406 Cohen and Rice 1995
3999 Dukes et al 1996
15629 Dukes et al 1995
15809 Oakley 1994
3592 Lindenberg et al 1998
13665 Rollin et al 1995
4024 Shope et al 1996
8339 Snow et al 1992
8113 Ellickson et al 1993
15907 Jason et al 1994
3189 Clapp and Early 1999
5874 Dykeman and Nelson 1996
4933 Donaldson et al 1994
4095 Goldberg et al 1996
4653 Botvin et al 1995
15971 Bekir et al 1993
12830 Halmi and Golik-Gruber 2001
3706 Fawcett et al 1997
8577 Wiggins 1991
15667 Delgado 1995
14044 Makkai et al 1991
13530 Hartman and Wolk 1996
13214 Sacks et al 1999
6527 Dembo et al 2002
4175 Frith 1996
8171 Boettcher 1993
13697 Schinke and Tepavac 1995
8701 Peine and Terry 1990
9126 Barret and White 1991
8073 Rickert et al 1993
16280 Caudill et al 1990
8337 Tobler 1992
5744 Coggan et al 1998
9260 Grossman et al 1993
4158 Waller et al 1996
8787 Graham et al 1990
8566 Hansen and Graham 1991

4908 Davis et al 1994
13945 Jenson et al 1993
8231 Becker et al 1992
8800 Wodak et al 1990
4248 Tricker and Connolly 1996
4070 Young et al 1996
8819 Klitzner et al 1990
8845 Moberg and Piper 1990
3943 Hansen and MacNeal 1997
4532 Cohen and Linton 1995
4569 Burns et al 1995
8931 Bell et al 1993
8521 Kreutter and Gewirtz 1991
4931 Ennett et al 1994
10729 Rohrbach et al 1994
8430 Richter et al 1991
4880 Ennett et al 1994
8448 Werch et al 1991
3165 Hall-Long and Dishop 1999
3514 Donnermeyer and Davis 1998
16077 Shope et al 1992
9282 Grossman and Gieck 1992
3023 Williams et al 1999
13314 Rosenbaum and Hanson 1998
13965 Manger et al 1992
13800 Felner et al 1994
8793 Jones et al 1990
7878 Corbin et al 1993
13446 Simpson 1997
8444 Leeds and D'Arcangelo (no date)
5865 Delgado 1997
8546 Parker et al 1991
4152 Hansen 1996
5816 Corby and Russell 1997
1555 Kumpfer et al 2002
10450 Mason 1996
1661 Werch and Owens 2002
14730 Collins et al 2002
13413 Laquer 1998
13412 Van Stelle et al 1998
606 LoSciuto and Steinman 2004
605 Abatemarco et al 2004
15969 Hecht et al 1993
13932 Botvin et al 1992
13931 Pentz et al 1992
3157 Bryson 1999
3159 Czuchry et al 1999
13127 Erhard 1999
3290 Dent et al 1998
3362 Shope et al 1998
3434 Brown and Kreft 1998
1726 Cuijpers 2002
5803 Perry and Williams 1998
13642 Harachi et al 1996
10181 Strader et al 1997
15534 Brown et al 1997
13894 Dishion 1993

14934 Eggert et al 2001
14932 Dishion and Kavanagh 2001
15559 Botvin (no date)
5310 Crosswaite et al 2004
5831 Wardet et al 1997
5889 Tudor-Smith et al 1995
15954 Lecca and Watts 1993
15613 Battistich et al 1996
15529 LoSciuto et al 1997
15466 Donaldson et al 1995
4958 Dorpat 1994
4957 Fleming 1994
8852 Ellickson and Bell 1990
15385 Hurry and McGurk 1997
16163 Black 1991
16159 Clayton et al 1991
16242 Edwards and Edwards 1990
8769 Klitzner et al 1990
8768 Kim et al 1990
8767 Bruvold 1990
8103 Ellickson et al 1993
15 Brown et al 2005

(352 papers)

Category (2): Costing or modelling of the impacts of substance misuse

1024 Caulkin et al 2004
2170 Chatterji et al 2001
14402 Swisher et al 2004

(3 papers)

Category (3): Effectiveness of intervention not established

15227 Dembo et al 1999
118 Griffin 2005
14791 Banwell et al 2002
14719 Maxwell 2002
14512 Allen et al 2003
307 Murray et al 2005
5162 Anderson et al 2005
14915 Griffin et al 2005
4051 Harding et al 1996
4057 Baldwin et al 1996
4197 Wasseff et al 1996
13515 Hritz and Gabow 1997
5025 Okamoto et al 2006
13477 Namyniuk 1997
634 Grace-Cleveland 2004
1025 Stockwell et al 2004
1202 Skara and Sussman 2003
14143 Greenberg et al 2005
2464 Tuttle et al 2000
2474 Conduct 2000
2786 Backett-Milburn and Wilson 2000
12346 Fishbein 2004

5548 Rollin et al 2000
5739 Reininger et al 1999
7960 Catalano et al 1993
5421 Allen 2003
5648 Paine-Andrews 2000
3776 Busen and Beech 1997
3661 Montoya et al 1997
4758 Stivers 1994
16715 Bond 1998
16688 Ansell 2001
2074 Fawcett et al 2001
5471 Ziewacz 2002
15265 Rose 1998
14658 Biglan and Smolkowski 2002
15050 Belgrave 2000
14881 St Pierre 2001
15188 Bronowski 1999
15498 Belgrave 1997
14576 Welsh 2003
14575 Paige 2003
15512 Shaw 1997
5752 Gance-Cleveland 1995
5370 Dusenbury et al 2003
5354 Case and Haines 2003
5339 McBride 2004
14227 Pratt 2004
14376 Anderson 2004
14458 Boohar 2003
14950 Clutterbuck 2001
14956 Maxwell 2000
14400 Baer et al 2004
5346 Patterson et al 2003
5560 Martino-McAllister et al 2001
14501 Accordino et al 2003
5814 Navarro et al 1997
4117 Farrell et al 1996
3648 Sussman et al 1997
14105 Shillington et al 2005
3303 Cheadle et al 1998
5884 Delgado 1995
5887 Mason and Hodge 1995
4988 Eggert et al 1994
4751 Carlson 1994
4989 Malekoff 1994
8695 McCarter 1990
14013 Schram et al 1991
4443 Elmquist 1995
13941 Bryce et al 1992
15656 Delgado 1996
8500 Miele 1991
8340 Starn 1992
8496 McGovern and DuPont 1991
15396 Miano et al 1997
3958 Sussman et al 1997
607 Martino-McAllister 2004
13575 Steneroth et al 1996
3219 Baldwin et al 1999

15630 Brown and Caston 1995
15527 Waters et al 1996
15525 James and Wabaunsee 1995
3430 D'Emidio-Caston and Brown 1998
3431 Rindskopf and Saxe 1998
5153 Peretti-Watel 2005
5830 Dawson 1997
4674 Hahn 1995
5800 O'Donnell et al 1997
15692 Aguirre-Molina and Parra 1995
13696 Palmgreen et al 1995
14764 Helitzer and Yoon 2002

(93 papers)

Category (4): 'Non-Systematic Reviews'

16658 Cuijpers 2003
15268 Newburn 1999
15092 Lloyd et al 2000
242 Velleman et al 2005
14199 Wolchik et al 2005
2855 Allott et al 1999
14991 Midford 2000
13732 Schydlower et al 1995
630 Sussman et al 2004
703 Tevyaw and Monti 2004
871 Kumpfer and Bluth 2004
13079 Gordon 2000
933 Evans-Whipp et al 2004
13078 Lochman 2000
1134 Stephenson 2003
1199 Kumpfer et al 2003
702 Wagner et al 2000
1238 Anderson and Loomis 2003
10131 Bosworth 1997
14144 Toumbourou et al 2005
9929 Sharp et al 1999
1574 Brown and Witherspoon 2002
1655 Spoth and Redmond 2002
1686 Andrews and Wilkinson (no date)
1705 Tubman et al 2002
2108 Webster-Stratton 2001
2398 Botvin and Kantor 2000
2435 Szapocznik and Williams 2000
2475 Dishion and Kavanagh 2000
2476 Botvin 2000
2631 Liddle and Hogue 2000
12144 Dowrick 2004
12177 Terzian 2005
12300 Botvin 2004
12608 Guerra 2003
12683 Schaps 2003
5540 Mathias 1997
5541 Wyman 1997
5574 Cohen and Rice 1996
3125 Anonymous 1999

5664 Plant and Plant 1999
5702 Kumpfer 1998
5708 Catalano 1998
12132 Larimer et al 2005
5921 Allensworth 1994
2836 Dishion et al 1999
9375 Walls 2003
9416 Kerka 2003
9451 McCoombs and Moore 2002
1245 Kumpfer and Alvarado 2003
1246 Nation et al 2003
1654 Kumpfer et al 2002
7998 Reniscow 1993
9430 Greenberg 2003
9437 Ruffolo 2003
9501 Liddle 2002
9521 Lochman 2002
9623 Blake 2001
9625 Guthrie 2001
9672 Beauvais 2001
9730 Kumpfer 2000
16710 Gilvarry 1998
5211 Smokowski et al 2005
1014 Hawkins et al 2004
8876 Maccoby 1990
14660 Botvin et al 2002
5270 Tayar et al 2004
5323 Campbell-Heider et al 2003
15414 Ashford 1997
13116 Albano 2000
5565 Shin 2001
5298 Pinto and Queely 2003
15960 Barth et al 1993
15953 Bukstein and Van Hasselt 1993
10547 Withers and Batten 1995
1540 Midford et al 2002
1230 Lilja et al 2003
14979 Hernandez et al 2000
559 De La Haye 2004
629 August et al 2004
5074 Burke et al 2005
3971 Hall and Zigler 1997
5846 Eggert et al 1996
15122 Evans 1999
9316 Felner et al 1991
9321 Logan 1991
1850 Botvin and Griffin 2002
5737 Schinke et al 1999
7874 Lamarine 1993
10040 Yuen and Nakano-Matsumoto 1998
8317 Botvin and Botvin 1992
8758 Schilling and McAlister 1990
8533 Pellow and Jengeleski 1991
8368 Brandon 1992
8047 Christopher et al 1993
5850 Davis 1997
8163 Wagner et al 1993

16762 Williams and Keene 1995
4979 Plotnick 1994
7924 Smith 1993
8409 O'Connor and Saunders 1992
10041 Yuen and Pardeck 1998
13990 England and Cole 1992
8801 DeJong and Winsten 1990
15777 Yanai and Weis 1994
8095 Hasselt et al 1993
15372 Cervantes and Pana 1998
15505 Beauvais 1997
8298 Perry and Kelder 1992
5730 Walsh 1997
2901 Hogue and Liddle 1999
15657 Blackman 1996
4357 Dusenbury and Falco 1995
5788 Korkia 1998
13449 Smyth and Saulnier 1996
5812 Orens et al 1997
2296 Vakalahi 2001
15952 Leventhal and Keeshan 1993
14181 Malekoff 2005
10234 Dusenbury and Falco 1997
5857 Schinke et al 1997
3432 Kreft 1998
5880 Perry 1996
15945 Santisteban et al 1993
16050 Chassin et al 1992
16028 Hill and Hill 1992
15545 Dembo and Rivers 1996
15336 Barret and Rafton 1998
16007 Greenwood 1992
14930 Gil et al 2001
14933 Bry and Attaway 2001
15862 Kelder and Perry 1993
14310 Calhoun et al 2005
14311 Glaser and Cohen 2005
15558 Kumpfer et al 2003
3433 Beck 1998
16239 Dryfoos 1990
14183 Swenson et al 2005
16241 Delgado 1990
16260 Williams 1990

(139 papers)

Appendix I

Categorization of 'Non-Systematic Reviews' Excluded from the Substance Misuse Review

Black and minority ethnic groups

1654 Kumpfer et al 2002
1014 Hawkins et al 2004
10040 Yuen and Nakano-Matsumoto 1998
15372 Cervantes and Pana 1998

Upstream, 'community' approaches

12683 Schaps 2003
9437 Ruffolo 2003
5857 Schinke et al 1997
16007 Greenwood 1992

Alternative conceptualizations of the phenomenon of substance misuse

16658 Cuijpers 2003
1230 Lilja et al 2003
8409 O'Connor and Saunders 1992
15657 Blackman 1996

Children of substance-using parents

5323 Campbell-Heider et al 2003
7924 Smith 1993

Substance users with mental health problems

16710 Gilvarry 1998

Family-based approaches

242 Velleman et al 2005
14199 Wolchik et al 2005
871 Kumpfer and Bluth 2004
13078 Lochman 2000
1199 Kumpfer et al 2003
2435 Szapocznik and Williams 2000
2475 Dishion and Kavanagh 2000
2631 Liddle and Hogue 2000
12177 Terzian 2005
5574 Cohen and Rice 1996
1245 Kumpfer and Alvarado 2003
9521 Lochman 2002
9730 Kumpfer 2000
5298 Pinto and Queely 2003
15122 Evans 1999
2901 Hogue and Liddle 1999
15558 Kumpfer et al 2003

Homeless

8163 Wagner et al 1993

Behaviour-modification or Life Skills Training

703 Tevyaw and Monti 2004
1705 Tubman et al 2002
12300 Botvin 2004
8047 Christopher et al 1993

Summaries of the broad policy consensus

1246 Nation et al 2003
1540 Midford et al 2002
4357 Dusenbury and Falco 1995

Methodological

8368 Brandon 1992
13449 Smyth and Saulnier 1996

Papers that would not add anything substantive to the review scope

15268 Newburn 1999
15092 Lloyd et al 2000
2855 Allott et al 1999
14991 Midford 2000
13732 Schydlower et al 1995
630 Sussman et al 2004
13079 Gordon 2000
933 Evans-Whipp et al 2004
1134 Stephenson 2003
702 Wagner et al 2000
1238 Anderson and Loomis 2003
10131 Bosworth 1997
9929 Sharp et al 1999
1574 Brown and Witherspoon 2002
1655 Spoth and Redmond 2002
1686 Andrews and Wilkinson (no date)
2108 Webster-Stratton 2001
2398 Botvin and Kantor 2000
2476 Botvin 2000
12144 Dowrick 2004
12608 Guerra 2003
5540 Mathias 1997
5541 Wyman 1997
3125 Anonymous 1999
5664 Plant and Plant 1999
5702 Kumpfer 1998
5708 Catalano 1998
12132 Larimer et al 2005
5921 Allensworth 1994
2836 Dishion et al 1999
9375 Walls 2003
9416 Kerka 2003
9451 McCoombs and Moore 2002

7998 Reniscow 1993
9430 Greenberg 2003
9501 Liddle 2002
9623 Blake 2001
9625 Guthrie 2001
9672 Beauvais 2001
5211 Smokowski et al 2005
5270 Tayar et al 2004
15414 Ashford 1997
5565 Shin 2001
629 August et al 2004
5074 Burke et al 2005
3971 Hall and Zigler 1997
7874 Lamarine 1993
8317 Botvin and Botvin 1992
5850 Davis 1997
16762 Williams and Keene 1995
4979 Plotnick 1994
10041 Yuen and Pardeck 1998
13990 England and Cole 1992
15777 Yanai and Weis 1994
15505 Beauvais 1997
8298 Perry and Kelder 1992
2296 Vakalahi 2001
3432 Kreft 1998
3433 Beck 1998

References

Abbott, S., Shaw, S. and Elston, J. (2004) 'Comparative analysis of health policy implementation'. *Policy Studies,* 25 (4), 259-66.

Acheson, D. (1998) *Independent Inquiry into Inequalities in Health Report.* London: The Stationery Office.

Aktan, G. B. (1999) 'A cultural consistency evaluation of a substance abuse prevention program with inner city African-American families'. *The Journal of Primary Prevention,* 19 (3), 227-239.

Arai, L., Britten, N., Popay, J., Roberts, H., Petticrew, M., Rodgers, M. and Sowden, A. (2007) 'Testing methodological developments in the conduct of narrative synthesis: a demonstration review of research on the implementation of smoke alarm interventions'. *Evidence & Policy,* 3 (3), 361-383.

Ashton, J. and Seymour, H. (1988) *The New Public Health: The Liverpool Experience.* Milton Keynes: Open University Press.

Asthana, S. and Halliday, J. (2006) 'Developing an evidence base for policies and interventions to address health inequalities: the analysis of 'Public Health Regimes''. *The Milbank Quarterly,* 84 (3), 577-603.

Atkinson, P. and Coffey, A. (2004) 'Analysing documentary realities'. In: D. Silverman (ed.), *Qualitative Research: Theory, Method and Practice.* London: Sage Publications, 56-75.

Bambra, C. L. (2005) 'Reviewing the evidence: reflections from experience'. *Evidence & Policy,* 1 (2), 243-255.

Baumler, A. (2001a) 'Introduction'. In: A. Baumler (ed.), *Modern China and Opium: A Reader.* Ann Arbor: University of Michigan Press, 1-5.

Baumler, A. (ed.) (2001b) *Modern China and Opium: A Reader.* Ann Arbor: University of Michigan Press.

Beaglehole, R., Bonita, R., Horton, R., Adams, O. and McKee, M. (2004) 'Public Health in the new era: Improving health through collective action'. *The Lancet,* 363 (9426), 2084-2087.

Bean, P. (1974) *The Social Control of Drugs.* London: Martin Robertson & Company.

Becker, H. (1963) *Outsiders: Studies in the Sociology of Deviance.* London: Free Press of Glencoe.

Bello, D. A. (2005) *Opium and the Limits of Empire: Drug Prohibition in the Chinese Interior, 1729-1850.* Cambridge, MA: Harvard University Press.

Berridge, V. (1980) 'The making of the Rolleston Report, 1908-1926'. *Journal of Drug Issues,* 10 (1), 7-28.

Berridge, V. (1984) 'Drugs and social policy: The establishment of drug control in Britain 1900-30'. *British Journal of Addiction,* 79 (1), 17-29.

Berridge, V. (1993) 'AIDS and British drug policy: continuity or change?' In: V. Berridge and P. Strong (eds.), *AIDS and Contemporary History*. Cambridge: Cambridge University Press, 135-156.

Berridge, V. (1996) *AIDS in the UK: The Making of Policy, 1981-1994*. Oxford: Oxford University Press.

Berridge, V. (1999) *Opium and the People: Opiate Use and Drug Control Policy in Nineteenth and Early Twentieth Century England*. London: Free Association Books.

Berridge, V. (2003) 'Public or policy understanding of history?' *Social History of Medicine*, 16 (3), 511-523.

Berridge, V. and Thom, B. (1996) 'Research and policy: what determines the relationship?' *Policy Studies*, 17 (1), 23-34.

Bhaskar, R. (1978) *A Realist Theory of Science*. Hemel Hempstead: Harvester Press.

Black, N. (2001) 'Evidence based policy: proceed with care'. *British Medical Journal*, 323 (7307), 275-279.

Blackman, S. (1996) 'Has drug culture become an inevitable part of youth culture? A critical assessment of drug education'. *Educational Review*, 48 (2), 131-142.

Blunden, C. and Elvin, M. (1983) *Cultural Atlas of China*. Oxford: Phaidon Press.

Boaz, A., Ashby, D., Denyer, D., Egan, M., Harden, A., Jones, D. R., Pawson, R. and Tranfield, D. (2006) 'A multitude of syntheses: a comparison of five approaches from diverse policy fields'. *Evidence & Policy*, 2 (4), 479-502.

Boaz, A. and Pawson, R. (2005) 'The perilous road from evidence to policy: five journeys compared'. *Journal of Social Policy*, 34 (2), 175-94.

Bonell, C., Oakley, A., Hargreaves, J., Strange, V. and Rees, R. (2006) 'Assessment of generalisability in trials of health interventions: Suggested framework and systematic review'. *British Medical Journal*, 333 (7563), 346-349.

Bosanquet, N. (1983) *After the New Right*. London: Heinemann Educational Books.

Bowker, G. C. and Star, S. L. (1999) *Sorting Things Out: Classification and its Consequences*. London: The MIT Press.

Boyle, J. S. (1994) 'Styles of ethnography'. In: J. M. Morse (ed.), *Critical Issues in Qualitative Research Methods*. London: Sage Publications, 159-185.

Bradshaw, P. L. and Bradshaw, G. (2004) *Health Policy for Health Care Professionals*. London: Sage Publications.

Braye, S. and Preston-Shoot, M. (2005) 'Emerging from out of the shadows? Service users and carer involvement in systematic reviews'. *Evidence & Policy*, 1 (2), 173-193.

Brewer, J. D. (2000) *Ethnography*. Buckingham: Open University Press.

Briggs, D. C. (2005) 'Meta-analysis: a case study'. *Evaluation Review*, 29 (2), 87-127.

British Sociological Association (2002) *Statement of Ethical Practice for the British Sociological Association*. Durham: British Sociological Association.

Broadbent, J. and Laughlin, R. (2002) 'Public service professionals and the New Public Management: control of the professions in the public services'. In: K. McLaughlin, S. P. Osborne and E. Ferlie (eds.), *New Public Management: Current Trends and Future Prospects*. London: Routledge, 95-108.

Brody, G. H., McBride Murry, V., Gerrard, M., Gibbons, F. X., Molgaard, V., McNair, L., Brown, A. C., Wills, T. A., Spoth, R. L., Luo, Z., Chen, Y. and Neubaum-Carlan, E. (2004) 'The Strong African American Families Program: Translating research into prevention programming'. *Child Development*, 75 (3), 900-917.

Brody, G. H., McBride Murry, V., McNair, L., Chen, Y., Gibbons, F. X., Gerrard, M. and Ashby Wills, T. (2005) 'Linking changes in parenting to parent-child relationship quality and youth self-control: The Strong African American Families Programme'. *Journal of Research on Adolescence*, 15 (1), 47-69.

Bruce, C. and Emshoff, J. (1992) 'The SUPER II program: An early intervention program'. *Journal of Community Psychology*, 20 (Special Issue), 10-21.

Brunton, G., Thomas, J., Harden, A., Rees, R., Kavanagh, J., Oliver, S., Shepherd, J. and Oakley, A. (2005) 'Promoting physical activity amongst children outside of physical education classes: a systematic review integrating intervention studies and qualitative studies'. *Health Education Journal*, 64 (4), 323-338.

Bruun, K., Pan, L. and Rexed, I. (1975) *The Gentlemen's Club: International Control of Drugs and Alcohol*. London: University of Chicago Press.

Bunge, M. (1998) *Philosophy of Science, Volume 1: From Problem to Theory*. London: Transaction Publishers.

Bunker, J. (2001) *Medicine Matters After All: Measuring the Benefits of Medical Care, a Healthy Lifestyle, and a Just Social Environment*. London: The Stationery Office.

Burr, A. (1987) 'Chasing the dragon: Heroin misuse, delinquency and crime in the context of South London culture'. *British Journal of Criminology*, 27 (4), 333-357.

Byrne, D. (2002) *Interpreting Quantitative Data*. London: Sage Publications.

Cabinet Office (1999) The Stationery Office, London.

Campbell, D. (1978) 'Qualitative knowing in Action Research'. In: P. Brenner, P. Marsh and M. Brenner (eds.), *The Social Contexts of Method*. London: Croom Helm, 184-209.

Campbell, D. (1986) 'Science's social system of validity-enhancing collective belief change and the problems of the social sciences'. In: D. W. Fiske and R. A. Shweder (eds.), *Metatheory in Social Science*. Chicago: University of Chicago Press, 108-135.

Campbell, D. and Russo, M. J. (1999) *Social Experimentation*. London: Sage Publications.

Campbell, D. and Stanley, J. (1963) *Experimental and Quasi-Experimental Evaluations in Social Research*. Chicago: Rand McNally.

Campbell, D., T (1988) 'Descriptive epistemology: Psychological, sociological, and evolutionary'. In: E. S. Overman (ed.), *Methodology and Epistemology for Social Science: Selected Papers*. London: University of Chicago Press, 435-486.

Canning, U., Millward, L., Raj, T. and Warm, D. (2004) *Drug Use Prevention Among Young People: A Review of Reviews. Evidence Briefing.* London: Health Development Agency.

Carr-Hill, R. (1995) 'Welcome? to the brave new world of evidence-based medicine'. *Social Science & Medicine,* 41 (11), 1467-1468.

Carstairs, C. (2005) 'The stages of the international drug control system'. *Drug and Alcohol Review,* 24 (1), 57-65.

Cervantes, R. C. and Pena, C. (1998) 'Evaluating Hispanic/Latino programs: Ensuring cultural competence'. *Alcoholism Treatment Quarterly,* 16 (1-2), 109-131.

Chalmers, A. F. (1999) *What is This Thing Called Science?* 3rd edition. Buckingham: Open University Press.

Chalmers, I. (2003) 'Trying to do more good than harm in policy and practice: the role of rigorous, transparent, up-to-date evaluations'. *Annals of the American Academy of Political and Social Science,* 589 (1), 22-40.

Chalmers, I. (2005) 'If evidence-informed policy works in practice, does it matter if it doesn't work in theory?' *Evidence & Policy,* 1 (2), 227-242.

Chalmers, I. and Altman, D. (eds.) (1995) *Systematic Reviews.* London: BMJ Publishing Group.

Chalmers, I., Enkin, M. and Keirse, M. J. N. C. (1989a) *Effective Care in Pregnancy and Childbirth. Volume 1: Pregnancy.* Oxford: Oxford University Press.

Chalmers, I., Enkin, M. and Keirse, M. J. N. C. (1989b) *Effective Care in Pregnancy and Childbirth. Volume 2: Childbirth.* Oxford: Oxford University Press.

Chalmers, I., Hedges, L. V. and Cooper, H. (2002) 'A brief history of research synthesis'. *Evaluation and the Health Professions,* 25 (1), 12-37.

Charlton, B. G. (1997) 'Restoring the balance: evidence-based medicine put in its place'. *Journal of Evaluation in Clinical Practice,* 3 (2), 87-98.

Chung, T. (1978) *China and the Brave New World: A Study of the Origins of the Opium War (1840-42).* Bombay: Allied Publishers Private Limited.

Clarence, E. (2002) 'Technocracy reinvented: the new evidence based policy movement'. *Public Policy and Administration,* 17 (3), 1-11.

Clarke, J. and Newman, J. (1997) *The Managerial State: Power, Politics and Ideology in the Remaking of Social Welfare.* London: Sage Publications.

Clarke, M. and Chalmers, I. (1998) 'Discussion sections in reports of controlled trials published in general medical journals: Islands in search of continents?' *Journal of the American Medical Association,* 280 (3), 280-282.

Clegg, S. (2005) 'Evidence-based practice in educational research: a critical realist critique of systematic review'. *British Journal of Sociology of Education,* 26 (3), 415-428.

Coggans, N., Cheyne, B. and McKellar, S. (2002) *The Life Skills Training Drug Education Programme: A Review of Research.* Edinburgh: Scottish Executive Drug Misuse Research Programme.

Considine, M. (2005) *Making Public Policy.* Cambridge: Polity Press.

Cook, T. D. and Campbell, D. (1979) *Quasi-Experimentation.* Chicago: Rand McNally.

Cooper, H. and Hedges, L. V. (1994) *The Handbook of Research Synthesis.* New York: Russell Sage.

Corti, L. (2004) 'Data archiving'. In: S. Becker and A. Bryan (eds.), *Understanding Research for Social Policy and Practice: Themes, Methods and Approaches.* Bristol: Policy Press, 148-152.

Courtwright, D. T. (2001) *Forces of Habit: Drugs and the Making of the Modern World.* Cambridge, MA: Harvard University Press.

Cuijpers, P. (2003) 'Three decades of drug prevention research'. *Drugs: Education, Prevention and Policy,* 10 (1), 7-20.

Davey Smith, G., Ebrahim, S. and Frankel, S. (2001) 'How policy informs the evidence - 'evidence based' thinking can lead to debased policy making'. *British Medical Journal,* 322 (7280), 184-185.

Davey Smith, G. and Egger, M. (1998) 'Incommunicable knowledge? Intrepreting and applying the results of clinical trials and meta-analyses'. *Journal of Clinical Epidemiology,* 51 (4), 289-295.

Davey Smith, G. and Egger, M. (2001) 'Going beyond the grand mean: subgroup analysis in meta-analysis of randomised trials'. In: M. Egger, G. Davey Smith and D. G. Altman (eds.), *Systematic Reviews in Health Care: Meta-Analysis in Context.* London: BMJ Books, 143-156.

Davies, H. and Nutley, S. (2000) 'Healthcare: evidence to the fore'. In: H. T. O. Davies, S. M. Nutley and P. C. Smith (eds.). Bristol: The Policy Press, 43-67.

Davies, H., Nutley, S. and Smith, P. (2000) 'Introducing evidence-based policy and practice in public services'. In: H. Davies, S. Nutley and P. Smith (eds.), *What Works? Evidence-Based Policy and Practice in Public Services.* Bristol: The Policy Press, 1-11.

Davies, J. B. (1997) *The Myth of Addiction.* 2nd edition. Amsterdam: Harwood Academic Publishers.

Deakin, N. and Walsh, K. (1996) 'The enabling state: The role of markets and contracts'. *Public Administration,* 74 (1), 33-48.

Dean, K. and Hunter, D. (1996) 'New directions for health: towards a knowledge base for public health action'. *Social Science & Medicine,* 42 (5), 745-751.

Department of Health (1992) *The Health of the Nation: A Strategy for Health in England.* London: HMSO.

Department of Health (1997) *The New NHS: Modern, Dependable. Cm 3807.* London: The Stationery Office.

Department of Health (1999) *Saving Lives: Our Healthier Nation. Cm4386.* London: The Stationery Office.

Department of Health (2001) *A Research and Development Strategy for Public Health.* London: Department of Health.

Department of Health and Social Security (1976) *Prevention and Health, Everybody's Business: A Reassessment of Public and Personal Health.* London: HMSO.

Dikotter, F., Laamann, L. and Xun, Z. (2004) *Narcotic Culture: A History of Drugs in China.* London: Hurst & Company.

Dixon-Woods, M., Agarwal, S., Young, B., Jones, D. and Sutton, A. (2004) *Integrative Approaches to Qualitative and Quantitative Evidence.* London: Health Development Agency.

Dixon-Woods, M., Bonas, S., Booth, A., Jones, D. R., Miller, T., Sutton, A. J., Shaw, R. L., Smith, J. A. and Young, B. (2006a) 'How can systematic reviews incorporate qualitative research? A critical perspective'. *Qualitative Research,* 6 (1), 27-44.

Dixon-Woods, M., Cavers, D., Agarwal, S., Annandale, E., Arthur, A., Harvey, J., Hsu, R., Katbamna, S., Olsen, R., Smith, L., Riley, R. and Sutton, A. J. (2006b) 'Conducting a critical interpretive synthesis of the literature on access to healthcare by vulnerable groups'. *BMC Medical Research Methodology,* 6 (35).

Dopson, S., Locock, L., Gabbay, J., Ferlie, E. and Fitzgerald, L. (2003) 'Evidence-based medicine and the implementation gap'. *Health,* 7 (3), 311-330.

Dorey, P. (2005a) 'Developments in British public policy'. In: P. Dorey (ed.), *Developments in British Public Policy.* London: Sage, 205-225.

Dorey, P. (2005b) *Policy Making in Britain: An Introduction.* London: Sage.

Dowding, K. (1995) 'Model or metaphor? A critical review of the policy network approach'. *Political Studies,* 43 (1), 136-158.

Duke, S. B. and Gross, A. C. (1993) *America's Longest War: Rethinking Our Tragic Crusade Against Drugs.* New York, NY: Jeremy P. Tarcher, Inc.

Eccles, M., Freemantle, N. and Mason, J. (2001) 'Using systematic reviews in clinical guideline development'. In: M. Egger, G. Davey Smith and D. G. Altman (eds.), *Systematic Reviews in Health Care: Meta-Analysis in Context.* London: BMJ Books, 400-409.

Edmonds, K., Sumnall, H., McVeigh, J. and Bellis, M. A. (2005) *Drug Prevention Among Vulnerable Young People.* Liverpool: National Collaborating Centre for Drug Prevention.

Egger, M. and Davey Smith, G. (2001) 'Principles and procedures for systematic reviews'. In: M. Egger, G. Davey Smith and D. Altman (eds.), *Systematic Reviews in Health Care: Meta-Analysis in Context.* London: BMJ Books, 23-42.

Egger, M., Davey Smith, G. and O'Rourke, K. (2001a) 'Rationale, potentials, and promise of systematic reviews'. In: M. Egger, G. Davey Smith and D. G. Altman (eds.), *Systematic Reviews in Health Care: Meta-Analysis in Context.* 2nd edition. London: BMJ Books, 3-19.

Egger, M., Dickersin, K. and Davey Smith, G. (2001b) 'Problems and limitations in conducting systematic reviews'. In: M. Egger, G. Davey Smith and D. G. Altman (eds.), *Systematic Reviews in Health Care: Meta-Analysis in Context*. 2nd edition. London: BMJ Books, 43-68.

Emerson, R. M., Fretz, R. I. and Shaw, L. L. (2001) 'Participant observation and fieldnotes'. In: P. Atkinson, A. Coffey, S. Delamont, J. Lofland and L. Lofland (eds.), *Handbook of Ethnography*. London: Sage Publications, 352-368.

Emshoff, J., Avery, E., Raduka, G., Anderson, D. J. and Calvert, C. (1996) 'Findings from SUPER STARS: A health promotion programme for families to enhance multiple protective factors'. *Journal of Adolescent Research,* 11 (1), 68-96.

Evidence-Based Medicine Working Group (1992) 'Evidence-Based Medicine: A new approach to teaching the practice of medicine'. *Journal of the American Medical Association,* 268 (17), 2420-2425.

Fairbank, J. K. (1978) 'The creation of the treaty system'. In: D. Twitchett and J. K. Fairbank (eds.), *The Cambridge History of China*. Cambridge: Cambridge University Press, 213-263.

Fairclough, N. (1992) *Discourse and Social Change*. Cambridge: Polity Press.

Fee, E. and Porter, D. (1992) 'Public health, preventive medicine and professionalisation: England and America in the nineteenth century'. In: A. Wear (ed.), *Medicine in Society: Historical Essays*. Cambridge: Cambridge University Press, 249-275.

Feinstein, A. R. and Horwitz, R. I. (1997) 'Problems in the 'evidence' of 'Evidence-Based Medicine''. *American Journal of Medicine,* 103 (6), 529-535.

Fischer, F. (2003) *Reframing Public Policy: Discursive Politics and Deliberative Practices*. Oxford: Oxford University Press.

Fischer, F. and Forester, J. (1993) 'Editors' introduction'. In: F. Fischer and J. Forester (eds.), *The Argumentative Turn in Policy Analysis and Planning*. London: University College London Press, 1-17.

Fitzgerald, L., Dopson, S., Ferlie, E. and Locock, L. (2005) 'Knowledge in action'. In: S. Dopson and L. Fitzgerald (eds.), *Knowledge to Action? Evidence-Based Health Care in Context*. Oxford: Oxford University Press, 155-181.

Foucault, M. (2002) *The Archaeology of Knowledge*. Original publication 1974. London: Routledge.

Freese, L. (1980) 'The problem of cumulative knowledge'. In: L. Freese (ed.), *Theoretical Methods in Sociology*. Pittsburgh, PA: University of Pittsburgh Press, 13-69.

Friedman, M. (2002) *Capitalism and Freedom*. Original publication 1962. Chicago: University of Chicago Press.

Gallagher, D. J. (2004) 'Educational research, philosophical orthodoxy and unfulfilled promises: the quandry of traditional research in US special education'. In: G. Thomas and R. Pring (eds.), *Evidence-Based Practice in Education*. Maidenhead: Open University Press, 119-130.

Gamble, A. (1986) 'The political economy of freedom'. In: R. Levitas (ed.), *Ideology and the New Right*. Cambridge: Polity Press, 25-54.

Ganguly, K. (2004) 'Opium use in Rajasthan, India: A socio-cultural perspective'. In: R. Coomber and N. South (eds.), *Drug Use and Cultural Contexts 'Beyond the West': Tradition, Change and Post-Colonialism*. London: Free Association Books, 83-100.

Glasgow, R. E., Green, L. W. and Ammerman, A. (2007) 'A focus on external validity'. *Evaluation and the Health Professions*, 30 (2), 115-117.

Glasziou, P. and Haynes, B. (2005) 'The paths from research to improved health outcomes'. *Evidence-Based Medicine*, 10 (1), 4-7.

Glasziou, P., Irwig, L., Bain, C. and Colditz, G. (2001) *Systematic Reviews in Health Care: A Practical Guide*. Cambridge: Cambridge University Press.

Glasziou, P., Vandenbroucke, J. and Chalmers, I. (2004) 'Assessing the quality of research'. *British Medical Journal*, 328 (7430), 39-41.

Godlee, F. and Dickersin, K. (2003) 'Bias, subjectivity, chance, and conflict of interest in editorial decisions'. In: F. Godlee and T. Jefferson (eds.), *Peer Review in Health Sciences*. 2nd edition. London: BMJ Books, 91-117.

Gomm, R., Hammersley, M. and Foster, P. (2000) 'Case study and generalization'. In: R. Gomm, M. Hammersley and P. Foster (eds.), *Case Study Method*. London: Sage, 98-115.

Goodman, J. (1995) 'Excitantia: Or, how enlightenment Europe took to soft drugs'. In: J. Goodman, P. E. Lovejoy and A. Sherratt (eds.), *Consuming Habits: Drugs in History and Anthropology*. London: Routledge, 126-147.

Goodman, N. (1978) *Ways of Worldmaking*. Hassocks: Harvester Press.

Gough, D. (2007) 'Weight of Evidence: a framework for the appraisal of the quality and relevance of evidence'. *Research Papers in Education*, 22 (2), 213-228.

Gough, D. and Elbourne, D. (2002) 'Systematic research synthesis to inform policy, practice and democratic debate'. *Social Policy & Society*, 1 (3), 225-236.

Graham, H. and Kelly, M. (2004) *Health Inequalities: Concepts, Frameworks and Policy*. London: Health Development Agency.

Gray, J. (1990) *Rebellions and Revolutions: China from the 1800s to 1980s*. Oxford: Oxford University Press.

Gray, J. (1993) *Beyond the New Right: Markets, Government and the Common Environment*. London: Routledge.

Gray, J. (1999) *False Dawn: The Delusions of Global Capitalism*. London: Granta Books.

Grayson, L. (2002) *Evidence based policy and the quality of evidence: rethinking peer review. ESRC UK Centre for Evidence Based Policy and Practice, Working Paper 7*. London: ESRC UK Centre for Evidence Based Policy and Practice.

Green, D. G. (1987) *The New Right: The Counter-Revolution in Political, Economic and Social Thought*. Brighton: Wheatsheaf Books.

Green, J. and Thorogood, N. (1998) *Analysing Health Policy: A Sociological Approach.* London: Longman.

Green, J. and Tones, K. (1999) 'Towards a secure evidence base for health promotion'. *Journal of Public Health Medicine,* 21 (2), 133-139.

Greenhalgh, T., Kristjansson, E. and Robinson, V. (2007) 'Realist review to understand the efficacy of school feeding programmes'. *British Medical Journal,* 335 (7625), 858-861.

Greenwood, P. W. (1992) 'Substance abuse problems among high-risk youth and potential interventions'. *Crime & Delinquency,* 38 (4), 444-458.

Griffin, K. W., Botvin, G. J. and Nichols, T. R. (2003) 'Effectiveness of a universal drug abuse prevention approach for youth at high risk for substance use initiation'. *Preventive Medicine,* 36 (1), 1-7.

Hajer, M. (1993) 'Discourse coalitions and the institutionalization of practice: the case of acid rain in Britain'. In: F. Fischer and J. Forester (eds.), *The Argumentative Turn in Policy Analysis and Planning.* London: University College London Press, 43-76.

Hajer, M. (2003) 'A frame in the fields: policymaking and the reinvention of politics'. In: M. Hajer and H. Wagenaar (eds.), *Deliberative Policy Analysis: Understanding Governance in the Network Society.* Cambridge: Cambridge University Press, 88-110.

Hajer, M. and Wagenaar, H. (2003) 'Introduction'. In: M. Hajer and H. Wagenaar (eds.), *Deliberative Policy Analysis: Understanding Governance in the Network Society.* Cambridge: Cambridge University Press, 1-30.

Hammersley, M. (1985) 'From ethnography to theory: a programme and a paradigm in the sociology of education'. *Sociology,* 19 (2), 244-259.

Hammersley, M. (1992) *What's Wrong with Ethnography?* London: Routledge.

Hammersley, M. (2001) 'On 'systematic' reviews of research literatures: a 'narrative' response to Evans and Benfield'. *British Educational Research Journal,* 27 (5), 543-554.

Hammersley, M. (2004) 'Some questions about evidence-based practice in education'. In: G. Thomas and R. Pring (eds.), *Evidence-Based Practice in Education.* Maidenhead: Open University Press, 133-149.

Hammersley, M. (2005) 'Is the evidence-based practice movement doing more good than harm? Reflections on Iain Chalmers' case for research-based policy making and practice'. *Evidence & Policy,* 1 (1), 85-100.

Hammersley, M. and Atkinson, P. (1995) *Ethnography: Principles in Practice.* 2nd edition. London: Routledge.

Harden, A. (2001) 'The fine detail: conducting a systematic review'. In: S. Oliver and G. Peersman (eds.), *Using Research for Effective Health Promotion.* Buckingham: Open University Press, 111-122.

Harding, G. (1988) *Opiate Addiction, Morality and Medicine: From Moral Illness to Pathological Disease.* Basingstoke: Macmillan Press.

Hawkins, E. H., Cummins, L. H. and Marlatt, G. A. (2004) 'Preventing substance abuse in American Indian and Alaska Native Youth: Promising strategies for healthier communities'. *Psychological Bulletin,* 130 (2), 304-323.

Haynes, B. (2005a) 'bmjupdates, a new FREE service for evidence-based clinical practice'. *Evidence-Based Medicine,* 10 (2), 35.

Haynes, B. (2005b) 'Of studies, summaries, synopses, and systems: the "4S" evolution of services for finding current best evidence'. *Evidence-Based Nursing,* 8 (1), 4-6.

Haynes, B. and Haines, A. (1998) 'Getting research findings into practice: barriers and bridges to evidence-based clinical practice'. *British Medical Journal,* 317 (7153), 273-276.

Health Development Agency (2004) *Public Health Evidence Steering Group. Second Annual Report.* London: Health Development Agency.

Hedges, L. V. and Olkin, I. (1985) *Statistical Methods for Meta-Analysis.* New York: Academic Press.

Henderson, S. (1997) *Ecstasy: Case Unsolved.* London: Pandora.

Higgins, J. P. T. and Green, S. (eds.) (2005) *Cochrane Handbook for Systematic Reviews of Interventions 4.2.5 (updated May 2005).* Chichester: John Wiley & Sons, Ltd.

Hill, M. (2000) *Understanding Social Policy.* 6th edition. Oxford: Blackwell.

Hogue, A., Liddle, H. A., Becker, D. and Johnson-Leckrone, J. (2002) 'Family-based prevention counseling for high-risk young adolescents: Immediate outcomes'. *Journal of Community Psychology,* 30 (1), 1-22.

Hogwood, B. and Gunn, L. (1984) *Policy Analysis for the Real World.* Oxford: Oxford University Press.

Holmes, D., Murray, S. J., Peron, A. and Rail, G. (2006) 'Deconstructing the evidence-based discourse in health sciences: truth, power and fascism'. *International Journal of Evidence Based Healthcare,* 4 (3), 180-186.

Home Office (1985) *Tackling Drug Misuse.* London: HMSO.

Hood, C. (1991) 'A New Public Management for all seasons?' *Public Administration,* 69 (1), 3-19.

Hood, C. (1995) 'Contemporary public management: A new global paradigm?' *Public Policy and Administration,* 10 (2), 104-117.

Howlett, M. and Ramesh, M. (2003) *Studying Public Policy: Policy Cycles and Policy Subsystems.* 2nd edition. Oxford: Oxford University Press.

Hugh-Jones, S. (1995) 'Coca, beer, cigars and yage: Meals and anti-meals in an Amerindian community'. In: J. Goodman, P. E. Lovejoy and A. Sherratt (eds.), *Consuming Habits: Drugs in History and Anthropology.* London: Routledge, 47-66.

Johnston, R. J., Gregory, D., Pratt, G. and Watts, M. (eds.) (2000) *The Dictionary of Human Geography.* 4th edition. Oxford: Blackwell.

Jones, H. (2005) 'Health policy'. In: P. Dorey (ed.), *Developments in British Public Policy*. London: Sage, 92-116.

Jones, L., Sumnall, H., Witty, K., Wareing, M., McVeigh, J. and Bellis, M. A. (2006) *A Review of Community-Based Interventions to Reduce Substance Misuse Among Vulnerable and Disadvantaged Young People*. Liverpool: National Collaborating Centre for Drug Prevention.

Jordan, A. G. and Richardson, J. J. (1987) *British Politics and the Policy Process: An Arena Approach*. London: Unwin Hyman.

Kelly, M. (2004) *The Evidence of Effectiveness of Public Health Interventions - and the Implications*. London: Health Development Agency.

Kelly, M. (2005) 'Public health guidance and the role of new NICE'. *Public Health*, 119 (11), 960-968.

Kelly, M. (2006a) 'The development of an evidence-based approach to tackling health inequalities in Britain'. In: A. Killoran, C. Swann and M. Kelly (eds.), *Public Health Evidence: Tackling Health Inequalities*. Oxford: Oxford University Press, 41-62.

Kelly, M. (2006b) 'Mapping the life-world: a future research priority for public health'. In: A. Killoran, C. Swann and M. Kelly (eds.), *Public Health Evidence: Tackling Health Inequalities*. Oxford: Oxford University Press, 553-573.

Kelly, M., McDaid, D., Ludbrook, A. and Powell, J. (2005) *Economic Appraisal of Public Health Interventions*. London: Health Development Agency.

Khan, K. S., ter Riet, G., Glanville, J., Sowden, A. J. and Kleijnen, J. (eds.) (2001) *Undertaking Systematic Reviews of Research on Effectiveness: CRD's Guidance for those carrying out or commissioning reviews*. York: University of York.

Killoran, A. and Kelly, M. (2004) 'Towards an evidence-based approach to tackling health inequalities: The English experience'. *Health Education Journal*, 63 (1), 7-14.

Knorr-Cetina, K. D. (1981) *The Manufacture of Knowledge: An Essay on the Constructivist and Contextual Nature of Science*. Oxford: Pergamon Press.

Kohn, M. (1992) *Dope Girls: The Birth of the British Drug Underground*. London: Lawrence & Wishart.

Kuhlmann, E. (2004) 'Knowledge and power in the era of consumerism - negotiations on professionalism in health care'. Paper for presentation at the International Sociological Association interim conference 'RC52, Knowledge, Work and Organization', 22-24 September, Versailles, www.zes.uni-bremen.de/~kuhlmann/ISA04_kuhlmann.pdf.

Kumpfer, K. L., Alvarado, R., Smith, P. and Bellamy, N. (2002) 'Cultural sensitivity and adaptation in family-based prevention interventions'. *Prevention Science*, 3 (3), 241-246.

Lancet (1994) 'Editorial: Population health looking upstream'. *Lancet*, 343 (8895), 429-430.

Latour, B. and Woolgar, S. (1986) *Laboratory Life: The Construction of Scientific Facts*. Princeton, NJ: Princeton University Press.

Legard, R., Keegan, J. and Ward, K. (2003) 'In-depth interviews'. In: J. Ritchie and J. Lewis (eds.), *Qualitative Research Practice: A Guide for Social Science Students and Researchers*. London: Sage Publications, 138-169.

Levine, H. G. and Reinarman, C. (1991) 'From prohibition to regulation: Lessons from alcohol policy for drug policy'. *The Milbank Quarterly*, 69 (3), 461-494.

Levitas, R. (1986) 'Ideology and the New Right'. In: R. Levitas (ed.), *The Ideology of the New Right*. Cambridge: Polity Press, 1-24.

Lewis, J. (1993) 'Public Health doctors and AIDS as a public health issue'. In: V. Berridge and P. Strong (eds.), *AIDS and Contemporary History*. Cambridge: University of Cambridge, 37-53.

Lilja, J., Larsson, S., Wilhelmsen, B. U. and Hamilton, D. (2003) 'Perspectives on preventing adolescent substance use and misuse'. *Substance Use & Misuse*, 38 (10), 1491-1530.

Lindblom, C. (1959) 'The science of 'muddling through''. *Public Administration Review*, 19 (2), 79-88.

Lindblom, C. (1979) 'Still muddling, not yet through'. *Public Administration Review*, 39 (6), 517-526.

Lipton, P. (2004) *Inference to the Best Explanation*. 2nd edition. London: Routledge.

Littlejohns, P. and Chalkidou, K. (2006) 'Evidence for health policy'. *Journal of Research in Nursing*, 11 (2), 110-117.

Longino, H. E. (1990) *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton, NJ: Princeton University Press.

Longino, H. E. (1996) 'Subjects, power and knowledge: Description and prescription in Feminist philosophies of science'. In: E. Fox Keller and H. E. Longino (eds.), *Feminism and Science*. Oxford: Oxford University Press, 264-279.

Majone, G. (1989) *Evidence, Argument and Persuasion in the Policy Process*. New Haven, Conn: Yale University Press.

Marshall, T. (1997) 'Scientific knowledge in medicine: a new clinical epistemology?' *Journal of Evaluation in Clinical Practice*, 3 (2), 133-138.

Mason, J. (2002) *Qualitative Researching*. 2nd edition. London: Sage Publications.

May, T. (2001) *Social Research: Issues, Methods and Process*. Maidenhead: Open University Press.

McAlister, F. (2001) 'Applying the results of systematic reviews at the bedside'. In: M. Egger, G. Davey Smith and D. G. Altman (eds.), *Systematic Reviews in Health Care: Meta-Analysis in Context*. London: BMJ Books, 373-385.

McCambridge, J. and Strang, J. (2004) 'The efficacy of single-session motivational interviewing in reducing drug consumption and perceptions of drug-related risk and harm among young people: results from a multi-site cluster randomized trial'. *Addiction*, 99 (1), 39-52.

McCambridge, J. and Strang, J. (2005) 'Deterioration over time in effect of motivational interviewing in reducing drug consumption and related risk among young people'. *Addiction,* 100 (4), 470-478.

McKeown, T. (1979) *The Role of Medicine: Dream, Mirage, or Nemesis?* 2nd edition. Oxford: Basil Blackwell.

Miles, A., Bentley, P., Polychronis, A., Grey, J. and Price, N. (1998) 'Recent progress in health services research: on the need for evidence-based debate'. *Journal of Evaluation in Clinical Practice,* 4 (4), 257-265.

Miles, A., Charlton, B., Bentley, P., Polychronis, A., Grey, J. and Price, N. (2000) 'New perspectives in the evidence-based healthcare debate'. *Journal of Evaluation in Clinical Practice,* 6 (2), 77-84.

Miller, G. (1997) 'Contextualizing Texts: Studying Organizational Texts'. In: G. Miller and R. Dingwall (eds.), *Context and Method in Qualitative Research.* London: Sage Publications, 77-91.

Miller, T., Bonas, S. and Dixon-Woods, M. (2007) 'Qualitative research on breastfeeding in the UK: a narrative review and methodological reflection'. *Evidence & Policy,* 3 (2), 197-230.

Milward, L. M., Kelly, M. and Nutbeam, D. (2003) *Public Health Interventions Research: The Evidence.* London: Health Development Agency.

Ministry of Health (1926) *Departmental Committee on Morphine and Heroin Addiction: Report.* London: HMSO.

Ministry of Health and Department of Health for Scotland (1961) *Drug Addiction. Report of the Interdepartmental Committee.* London: HMSO.

Ministry of Health and Scottish Home and Health Department (1965) *Drug Addiction. The Second Report of the Interdepartmental Committee.* London: HMSO.

Morse, J. M. and Richards, L. (2002) *Readme First for a User's Guide to Qualitative Methods.* London: Sage Publications.

Mott, J. and Bean, P. (1998) 'The development of drug control in Britain'. In: R. Coomber (ed.), *The Control of Drugs and Drug Users: Reason or Reaction?* Amsterdam: Harwood Academic Publishers, 31-48.

Muir Gray, J. A. (2001) 'Using systematic reviews for evidence based policy making'. In: M. Egger, G. Davey Smith and D. G. Altman (eds.), *Systematic Reviews in Health Care: Meta-Analysis in Context.* London: BMJ Books, 410-418.

Muir Gray, J. A. (2005) 'Evidence-based and value-based healthcare'. *Evidence-Based Healthcare & Public Health,* 9 (5), 317-318.

Mulkay, M. (1991) *Sociology of Science: A Sociological Pilgrimage.* Milton Keynes: Open University Press.

Mulrow, C. D. (1987) 'The medical review article: state of the science'. *Annals of Internal Medicine,* 106 (3), 485-488.

Murphy, M. K., Black, N. A., Lamping, D. L., McKee, C. M., Sanderson, C. F. B., Askham, J. and Marteau, T. (1998) 'Consensus development methods and their use in clinical guideline development'. *Health Technology Assessment,* 2 (3), 1-88.

Musto, D. F. (1999) *The American Disease: Origins of Narcotic Control.* 3rd edition. Oxford: Oxford University Press.

Nagel, E. (1979) *The Structure of Science: Problems in the Logic of Scientific Explanation.* London: Routledge and Kegan Paul.

National Institute for Health and Clinical Excellence (2006a) *Involving Patients and the Public in NICE guidance.* London: National Institute for Health and Clinical Excellence.

National Institute for Health and Clinical Excellence (2006b) *Methods for Development of NICE Public Health Guidance.* London: National Institute for Health and Clinical Excellence.

National Institute for Health and Clinical Excellence (2007) *Community-Based Interventions to Reduce Substance Misuse Among Vulnerable and Disadvantaged Children and Young People.* London: National Institute for Health and Clinical Excellence [Available online: http://www.nice.org.uk/PHI4].

Newman, J. (2000) 'Beyond the new public management? Modernising public services'. In: J. Clarke, S. Gewirtz and E. McLaughlin (eds.), *New Managerialism, New Welfare?* London: Sage Publications, 45-61.

Newman, J. (2002) 'The New Public Management, modernization and institutional change: disruptions, disjunctures and dilemmas'. In: K. McLaughlin, S. P. Osborne and E. Ferlie (eds.), *New Public Management: Current Trends and Future Prospects.* London: Routledge, 77-91.

Newman, R. K. (1995) 'Opium smoking in late imperial China: A reconsideration'. *Modern Asian Studies,* 29 (4), 765-794.

Noblit, G. W. and Hare, R. D. (1988) *Meta-Ethnography: Synthesizing Qualitative Studies.* London: Sage Publications.

Nutley, S., Walter, I. and Davies, H. (2002) *From Knowing to Doing: A Framework for Understanding the Evidence-Into-Practice Agenda. Discussion Paper 1.* St. Andrews: University of St. Andrews, Research Unit for Research Utilisation.

Oakley, A. (2000) *Experiments in Knowing: Gender and Method in Social Sciences.* Cambridge: Polity Press.

Oakley, A. (2006) 'Resistances to 'new' technologies of evaluation: Education research in the UK as a case study'. *Evidence & Policy,* 2 (1), 63-87.

Oakley, A., Gough, D., Oliver, S. and Thomas, J. (2005) 'The politics of evidence and methodology: lessons from the EPPI-centre'. *Evidence & Policy,* 1 (1), 5-31.

Oakley, A., Strange, V., Stephenson, J., Forrest, S. and Monteiro, H. (2004) 'Evaluating processes: a case study of a randomized controlled trial of sex education'. *Evaluation,* 10 (4), 440-462.

Oakley, A., Strange, V., Toroyan, T., Wiggins, M., Roberts, I. and Stephenson, J. (2003) 'Using random allocation to evaluate social interventions: Three recent UK

examples'. *Annals of the American Academy of Political and Social Science,* 589 (1), 170-189.

O'Connor, J. and Saunders, B. (1992) 'Drug education - An appraisal of a popular preventive'. *International Journal of the Addictions,* 27 (2), 165-185.

Oliver, S., Harden, A., Rees, R., Shepherd, J., Brunton, G., Garcia, J. and Oakley, A. (2005) 'An emerging framework for including different types of evidence in systematic reviews for public policy'. *Evaluation,* 11 (4), 428-446.

Osborne, D. and Gaebler, T. (1992) *Reinventing Government: How the Entrepreneurial Spirit is Transforming the Public Sector.* Wokingham: Addison-Wesley.

Osborne, S. P. and McLaughlin, K. (2002) 'The New Public Management in context'. In: K. McLaughlin, S. P. Osborne and E. Ferlie (eds.), *New Public Management: Current Trends and Future Prospects.* London: Routledge, 7-14.

Oxman, A. D. and Guyatt, G. H. (1988) 'Guidelines for reading literature reviews'. *Canadian Medical Association Journal,* 138 (8), 697-703.

Pagliari, C., Grimshaw, J. and Eccles, M. (2001) 'The potential influence of small group processes on guideline development'. *Journal of Evaluation in Clinical Practice,* 7 (2), 165-173.

Parsons, W. (2002) 'From muddling through to muddling up: Evidence based policy making and the modernisation of British Government'. *Public Policy and Administration,* 17 (3), 43-60.

Parssinen, T. M. and Kerner, K. (1980) 'Development of the disease model of drug addiction in Britain 1870-1926'. *Medical History,* 24 (3), 275-296.

Pawson, R. (2002) 'Evidence-based policy: the promise of 'realist synthesis''. *Evaluation,* 8 (3), 340-358.

Pawson, R. (2006) *Evidence-Based Policy: A Realist Perspective.* London: Sage Publications.

Pawson, R. and Tilley, N. (1997) *Realistic Evaluation.* London: Sage Publications.

Pearson, G. (1987) 'Social deprivation, unemployment and patterns of heroin use'. In: N. Dorn and N. South (eds.), *A Land Fit for Heroin? Drug Policies, Prevention and Practice.* Basingstoke: Macmillan Education, 62-94.

Petticrew, M. and Roberts, H. (2006) *Systematic Reviews in the Social Sciences: A Practical Guide.* Oxford: Blackwell.

Pietroni, P., Winkler, F. and Graham, L. (2003) 'Cultural revolution'. *British Medical Journal,* 326 (7404), 1304-1306.

Pollitt, C. (1990) *Managerialism and the Public Services: The Anglo-American Experience.* Oxford: Basil Blackwell.

Pollitt, C. (2003) *The Essential Public Manager.* Maidenhead: Open University Press.

Polychronis, A., Miles, A. and Bentley, P. (1996) 'Evidence-Based Medicine: Reference? Dogma? Neologism? New orthodoxy?' *Journal of Evaluation in Clinical Practice,* 2 (1), 1-3.

Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M. and Britten, N. (2006) *Guidance on the Conduct of Narrative Synthesis in Systematic Reviews: A Product from the ESRC Methods Programme.*

Pope, C. (2003) 'Resisting evidence: The study of evidence-based medicine as a contemporary social movement'. *Health,* 7 (3), 267-282.

Pope, C., Mays, N. and Popay, J. (2007) *Synthesizing Qualitative and Quantitative Health Evidence: A Guide to Methods.* Maidenhead: Open University Press.

Popper, K. (1972) *Objective Knowledge.* Oxford: Clarendon Press.

Potter, J. and Wetherell, M. (1987) *Discourse and Social Psychology.* London: Sage Publications.

Power, M. (1997) *The Audit Society: Rituals of Verification.* Oxford: Oxford University Press.

Prior, L. (2003) *Using Documents in Social Research.* London: Sage Publications.

Rhodes, R. A. W. (1997) *Understanding Governance: Policy Networks, Governance, Reflexivity and Accountability.* Buckingham: Open University Press.

Rhodes, T., Singer, M., Bourgois, P., Friedman, S. R. and Strathdee, S. A. (2005) 'The social structural production of HIV risk among injecting drug users'. *Social Science & Medicine,* 61 (5), 1026-1044.

Ritchie, J., Lewis, J. and Elam, G. (2003) 'Designing and selecting samples'. In: J. Ritchie and J. Lewis (eds.), *Qualitative Research Practice: A Guide for Social Science Students and Researchers.* London: Sage Publications, 77-108.

Rodner, S. (2005) '"I am not a drug abuser, I am a drug user": A discourse analysis of 44 drug users' consruction of identity'. *Addiction Research and Theory,* 13 (4), 333-46.

Rose, G. (1992) *The Strategy of Preventive Medicine.* Oxford: Oxford University Press.

Ruffolo, M. C., Evans, M. E. and Lukens, E. P. (2003) 'Primary prevention programs for children in the social service system'. *The Journal of Primary Prevention,* 23 (4), 425-450.

Rycroft-Malone, J. (2006) 'The politics of the evidence-based practice movements'. *Journal of Research in Nursing,* 11 (2), 95-108.

Sackett, D. L., Richardson, W. S., Rosenberg, W. and Haynes, R. B. (1997) *Evidence-Based Medicine: How to Practice and Teach EBM.* London: Churchill Livingstone.

Sackett, D. L., Rosenberg, W. M. C., Gray, J. A. M., Haynes, R. B. and Richardson, W. S. (1996) 'Evidence-based medicine: what it is and what it isn't'. *British Medical Journal,* 312 (7023), 71-72.

Sackett, D. L. and Wennberg, J. E. (1997) 'Choosing the best research design for each question: it's time to stop squabbling over the 'best' methods'. *British Medical Journal,* 315 (7123), 1636.

Saltz, R. F. (2005) 'What is 'evidence' and can we provide it?' In: T. Stockwell, P. J. Gruenewald, J. W. Toumbourou and W. Loxley (eds.), *Preventing harmful substance*

use: The evidence base for policy and practice. Chichester: John Wiley & Sons, Ltd, 313-324.

Sanderson, I. (2002) 'Making sense of 'what works': Evidence based policy making as instrumental rationality?' *Public Policy and Administration,* 17 (3), 61-75.

Sayer, A. (1992) *Method in Social Science.* 2nd edition. London: Routledge.

Sayer, A. (2000) *Realism and Social Science.* London: Sage.

Schaps, E. and Solomon, D. (2003) 'The role of the school's social environment in preventing student drug use'. *The Journal of Primary Prevention,* 23 (3), 299-328.

Schinke, S., Cole, K., Diaz, T. and Botvin, G. J. (1997) 'Developing and implementing interventions in community settings'. *Journal of Child & Adolescent Substance Abuse,* 6 (1), 49-67.

Schofield, J. W. (2000) 'Increasing the generalisability of qualitative research'. In: R. Gomm, M. Hammersely and P. Foster (eds.), *Case Study Method.* London: Sage, 69-97.

Seale, C. (1998) 'Qualitative interviewing'. In: C. Seale (ed.), *Researching Society and Culture.* London: Sage Publications.

Shadish, W. R., Cook, T. D. and Campbell, D. (2002) *Experimental and Quasi-Experimental Designs for Generalized Causal Inference.* Boston: Houghton Mifflin Company.

Sharland, E. and Taylor, I. (2006) 'Social care research: a suitable case for systematic review?' *Evidence & Policy,* 2 (4), 503-523.

Shepherd, J., Harden, A., Rees, R., Brunton, G., Garcia, J., Oliver, S. and Oakley, A. (2006) 'Young people and healthy eating: a systematic review of research on barriers and facilitators'. *Health Education Research,* 21 (2), 239-257.

Sherratt, A. (1995) 'Introduction: Peculiar Substances'. In: J. Goodman, P. E. Lovejoy and A. Sherratt (eds.), *Consuming Habits: Drugs in History and Anthropology.* London: Routledge, 1-10.

Silverman, D. (2001) *Interpreting Qualitative Data: Methods for Analysing Talk, Text and Interaction.* 2nd edition. London: Sage Publications.

Silverman, D. (2005) *Doing Qualitative Research: A Practical Handbook.* 2nd edition. London: Sage Publications.

Simon, H. A. (1957) *Administrative Behaviour.* 2nd edition. New York: Macmillan.

Slavin, R. E. (1995) 'Best evidence synthesis: an intelligent alternative to meta-analysis'. *Journal of Clinical Epidemiology,* 48 (1), 9-18.

Smart, C. (1984) 'Social policy and drug addiction: a critical study of policy development'. *British Journal of Addiction,* 79 (1), 31-39.

Smith, A. F. M. (1996) 'Mad cows and Ecstasy: chance and choice in an evidence-based society'. *Journal of the Royal Statistical Society Series A (Statistics in Society),* 159 (3), 367-383.

Smith, E. A., Swisher, J. D. and Vicary, J. R. (2004) 'Evaluation of life skills training and infused-life skills training in a rural setting: Outcomes at two years'. *Journal of Alcohol and Drug Education*, 48 (1), 51-70.

Solesbury, W. (2001) *Evidence Based Policy: Whence it Came and Where it's Going. ESRC UK Centre for Evidence Based Policy and Practice Working Paper 1.* London: ESRC UK Centre for Evidence Based Policy and Practice.

South, N. (1998) 'Tackling drug control in Britain: from Sir Malcolm Delevingne to the New Drugs Strategy'. In: R. Coomber (ed.), *The Control of Drugs and Drug Users: Reason or Reaction?* Amsterdam: Harwood Academic Publishers, 87-106.

Spear, B. (2005) 'The early years of Britain's drug situation in practice: Up to the 1960s'. In: J. Strang and M. Gossop (eds.), *Heroin Addiction and the British System Volume 1: Origins and Evolution.* London: Routledge, 17-42.

Spear, H. B. (1969) 'The growth of heroin addiction in the United Kingdom'. *British Journal of Addiction*, 64 (2), 245-255.

Spencer, L., Ritchie, J. and O'Connor, W. (2003) 'Analysis: practices, principles and processes'. In: J. Ritchie and J. Lewis (eds.), *Qualitative Research Practice: A Guide for Social Science Students and Researchers.* London: Sage, 199-218.

Spoth, R., Guyll, M., Chao, W. and Molgaard, V. (2003) 'Exploratory study of a preventive intervention with general population African American families'. *Journal of Early Adolescence*, 23 (4), 435-468.

Stake, R. E. (1995) *The Art of Case Study Research.* London: Sage.

Stimson, G. (1987) 'The war on heroin: British policy and the international trade in illicit drugs'. In: N. Dorn and N. South (eds.), *A Land Fit for Heroin? Drug Policies, Prevention and Practice.* Basingstoke: Macmillan Education, 35-61.

Stimson, G. (1990) 'Revising policy and practice: New ideas about the drug problem'. In: J. Strang and G. Stimson (eds.), *AIDS and Drug Misuse.* London: Routledge, 121-131.

Swann, C., Falce, C., Morgan, A., Kelly, M. and Powell, G. (2003) *Health Development Agency Evidence Base: Process and Quality Standards Manual for Evidence Briefings.* London: Health Development Agency.

Tannenbaum, S. J. (1993) 'What physicians know'. *New England Journal of Medicine*, 329 (17), 1268-1271.

Taylor, A. (1993) *Women Drug Users: An Ethnography of a Female Injecting Community.* Oxford: Clarendon Press.

Thom, B. (2005) 'Who makes alcohol policy? Science and policy networks, 1950-2000'. In: V. Berridge (ed.), *Making Health Policy: Networks in Research and Policy After 1945.* Amsterdam: Rodopi B.V, 75-99.

Throgmorton, J. A. (1993) 'Survey research as rhetorical trope: electric power planning arguments in Chicago'. In: F. Fischer and J. Forester (eds.), *The Argumentative Turn in Policy Analysis and Planning.* London: University College London Press, 117-143.

Townsend, P., Davidson, N. and Whitehead, M. (eds.) (1992) *Inequalities in Health: The Black Report and The Health Divide.* 2nd edition. Harmondsworth: Penguin Books.

Trocki, C. A. (1999) *Opium, Empire and the Global Political Economy: A Study of the Asian Opium Trade 1750-1950.* London: Routledge.

Vicary, J. R., Henry, K. L. and Bechtel, L. J. (2004) 'Life skills training effects for high and low risk rural junior high school females'. *Journal of Primary Prevention,* 25 (4), 399-416.

Victora, C. G., Habicht, J.-P. and Bryce, J. (2004) 'Evidence-Based Public Health: moving beyond randomized trials'. *American Journal of Public Health,* 94 (3), 400-408.

Wallace, A., Croucher, K., Bevan, M., Jackson, K., O'Malley, L. and Quilgars, D. (2006) 'Evidence for policy making: some reflections on the application of systematic reviews to housing research'. *Housing Studies,* 21 (2), 297-314.

Walsh, D. (1998) 'Doing Ethnography'. In: C. Seale (ed.), *Researching Society and Culture.* London: Sage Publications, 217-232.

Wanless, D. (2004) *Securing Good Health for the Whole Population: Final Report.* London: HMSO.

Weightman, A., Ellis, S., Cullum, A., Sander, L. and Turley, R. (2005) *Grading Evidence and Recommendations for Public Health Interventions: Developing and Piloting a Framework.* London: Health Development Agency.

Weiss, C. H. (1979) 'The many meanings of research utilization'. *Public Administration Review,* 39 (5), 426-431.

Weiss, C. H. (1987) 'The circuitry of enlightenment: Diffusion of Social Science research to policymakers'. *Knowledge: Creation, Diffusion, Utilization,* 8 (2), 274-281.

Wetherell, M., Taylor, S. and Yates, S. J. (2001) *Discourse Theory and Practice.* London: Sage Publications.

Whitehead, M., Diderichsen, F. and Burstrom, B. (2000) 'Researching the impact of public policy on inequalities in health'. In: H. Graham (ed.), *Understanding Health Inequalities.* Buckingham: Open University Press, 203-218.

Williams, M. (2005) 'Situated Objectivity'. *Journal for the Theory of Social Behaviour,* 35 (1), 99-120.

Williams, M. and May, T. (1996) *Introduction to the Philosophy of Social Research.* Abingdon: Routledge.

Wolcott, H. F. (1995) *The Art of Fieldwork.* London: Altamira Press.

Worrall, J. (2002) '*What* evidence in Evidence-Based Medicine?' *Philosophy of Science,* 69 (3), 316-330.

Yates, R. (2002) 'A brief history of British drug policy 1950-2001'. *Drugs: Education, Prevention and Policy,* 9 (2), 113-124.

Yin, R. K. (2003) *Case Study Research: Design and Methods.* 3rd edition. London: Sage Publications.

Young, K., Ashby, D., Boaz, A. and Grayson, L. (2002) 'Social Science and the Evidence-Based policy movement'. *Social Policy & Society,* 1 (3), 215-224.

Yuen, F. K. O. and Nakano-Matsumoto, N. (1998) 'Effective substance abuse treatment for Asian American adolescents'. *Early Child Development and Care,* 147 (1), 43-54.

Zinberg, N. E. (1984) *Drug, Set and Setting: The Basis for Controlled Intoxicant Use.* New Haven, Conn: Yale University Press.

Publications arising from this thesis

1) Pearson, M. (2007) Systematic Reviews in Social Policy: To go forward, do we first need to look back? *Evidence & Policy*, 3 (4), 505-526

2) Pearson, M. (2008, forthcoming) Evidence-Based Policy & Practice: How can research be made more relevant to practitioners and policy-makers? In: M.Wright (ed.) *Proceedings of the Faculty of Social Science & Business (University of Plymouth) Postgraduate Symposium 2007.* Plymouth: University of Plymouth

Mark Pearson, School of Law & Social Science, 15 Portland Villas, University of Plymouth. PL4 8AA.

Evidence-Based Policy & Practice: How can research be made more relevant to practitioners and policy-makers?

Abstract

Evidence-Based Policy and Practice has rapidly grown in an environment where a key New Labour policy has been to establish 'what works'. This growth marks an important development in how practitioners and policy-makers work – if decisions are not based upon 'the best evidence', then what are they based upon? However, the utilisation of research knowledge has proven to be far from straightforward; it is contested, contradictory, and difficult to apply in local contexts. It is perhaps not surprising that the 'hierarchy of evidence', which arbitrates between what is and is not acceptable as evidence on the basis of methodological criteria, holds such appeal.

This paper considers how research in the social arena can be made more useful to, and thus more likely to be used by, practitioners and policy-makers. It is not proposed that research should be 'policy-led', for that suggests an unquestioning adherence to the latest political buzzwords and a neglect of the critical function that is so important in academic research. Rather, the work of a key social research methodologist (Donald Campbell) on research validity is drawn upon. Through a consideration of how internal, external, and construct validity can be attained in social research, the paper explores how social research may be made both more relevant to, and more constructively critical of, policy and practice.

How should research inform professional practice? What role should research play in the policy-making process? The rise of Evidence-Based Policy & Practice (EBP) has raised the status of research; what practitioners do with their clients, and how policy shapes that practice and the institutional frameworks within which they occur, should be based upon the best possible evidence (Chalmers and Altman, 1995; Chalmers, 2003; Oakley et al., 2005). Such a statement may appear uncontroversial, even obvious or banal, but debate over what may qualify as evidence, and what weight it should be given in decision-making, is extensive and at times very heated (Chalmers, 2005; Hammersley, 2005).

This paper begins with a brief introduction to the rationale for EBP and the genesis of the debates within it; however, the bulk of the paper considers how researchers could make their research more applicable to the needs of practitioners and policy-makers so that they may make better use of this knowledge. I argue that, rather than the usual approach of asking, "How may we facilitate practitioners to *implement* these research findings?" (see, for

example, Haynes and Haines, 1998; Glasziou and Haynes, 2005), social policy and practice might be better informed by taking a further step back in order to consider *how* practitioners and policy-makers endeavour to utilise research findings. I contend that the perception that these parties work in a mechanistic fashion ("Here are the best research findings, therefore we should do what they say") is largely mistaken (see Holmes et al. (2006) for an extreme example of this); both practitioners and policy-makers bring far more, in the form of their professional experience and critical faculties, to the point at which they make decisions.

In view of the nature of this decision-making, I argue that what practitioners and policy-makers need are not fully formed answers to problems, but valid research that may be judiciously applied across a range of contexts outside of the particulars of the primary research itself. I contend that a promising, but under-explored, way to achieve this may be found in the work of Donald Campbell and colleagues (Campbell and Stanley, 1963; Campbell, 1978; Cook and Campbell, 1979; Campbell and Russo, 1999; Shadish et al., 2002). Campbell wrote extensively, and often in a manner that was not particularly accessible, on how research could best inform 'the experimenting society'[1] so that pressing social problems could be addressed using the best possible evidence. This paper focuses upon the different types of validity within research and the manner in which they are key to understanding research design. I argue that an understanding of internal, external and construct validity (see, in particular, Shadish et al., 2002, p.53-92) provides the building blocks (or perhaps 'thinking blocks' would be a better term) for designing and conducting research that is more likely to be of use to practitioners and policy-makers.

A brief history of Evidence-Based Policy & Practice

Three main factors can be argued to have converged in such a way as to give rise to EBP within the United Kingdom. Firstly, there was the recognition that many areas of social policy have a burgeoning literature that it is impossible for even a skilled and dedicated professional to keep abreast of; if decisions are to be made that take full account of the breadth of research and its findings (both positive and negative), there need to be methods to cumulate and synthesise this knowledge (Chalmers, 2003), and the institutional frameworks in which to do so (Oakley, 2006). Secondly, with the development of computer technology, it became possible to search very large databases comprehensively and obtain copies of relevant papers within a realistic timescale. Thirdly, the wider cultural development within the UK that saw the 'decline from priesthood' of the professions and a significant rise in the status of users of public services from passive recipients to knowledgeable 'consumers' (Gabe et al., 1994; Williams and Calnan, 1996; Pietroni et al., 2003), required that knowledge be made openly accessible in order that the power of experts could be challenged. It has led to the current situation where both government and public service professions may be held accountable for making decisions based upon the best evidence (Cabinet Office, 1999; Department of Health, 1999; Home Office, 1999). At the very

least, it is intended that a greater onus be placed on explicit decision-making based upon 'evidence' rather than relying upon opaque 'expertise'.

Whilst EBP has effected a change in the way that policy-making occurs, it has also built upon existing professional power structures, their inherent methodological preferences and policy histories. EBP with respect to social policy is strongly rooted in Evidence-Based Medicine (EBM), and in particular to a 'hierarchy of evidence' which follows a gradation from randomised-controlled trials (strong evidence), through cohort and case-control studies (not such good evidence), and down to expert opinion (weak evidence)[2]. This hierarchy may indeed be wholly justifiable when cumulating research conducted in tightly-controlled contexts where the principles underlying the phenomena being investigated are well understood. Whether this model is suitable for utilisation in areas of social policy, where experimental control presents substantive problems (Tilley, 2001; Byrne, 2002; Victora et al., 2004) and where our understanding of the mechanisms of social phenomena arguably lags significantly behind Medicine's understanding of the physiology of the human body (Kelly, 2006), is a moot point.

Sources of validity: Internal, External, and Construct

I do not propose to enter here into the debate about the utility of the hierarchy of evidence within social policy; the interested reader can follow critiques of EBM within medicine itself (Tannenbaum, 1993; Polychronis et al., 1996; Goodman, 1999) and also within health and social policy (Hammersley, 2004; Victora et al., 2004; Clegg, 2005; Oakley et al., 2005; Oakley, 2006) if they so wish. Nor is this the place for an analysis of a particular policy history. Instead, I want to take a step further back and propose how researchers could make their research of greater utility for practitioners and policy-makers, for whatever the pros and cons of the manner in which EBP has been implemented in the UK, at its heart lies a commitment to utilising research as a form of evidence. I propose that giving consideration to the three aspects of research validity, as defined by Donald Campbell and colleagues (Shadish et al., 2002), would provide a sound basis for researchers to design and conduct their research so that it better addresses the requirements of practitioners and policy-makers[3]. The three aspects are as follows:

1. Internal validity:- "The validity of inferences about whether the relationship between two variables is causal." (Shadish et al., 2002, p.508)

This refers to how sure one may be that, with the variables operationalised as decided upon by the researchers, the study was conducted in such a way that one can be sure that intervention x produces a change in phenomenon y. The essence of the approach is experimental isolation – the two factors that are of interest are isolated so that one can be certain that it can only be intervention x producing the change in phenomenon y. Campbell and Stanley's (1963) development of the means by which threats to internal validity could be addressed is clear in the hierarchy of evidence commonly utilised in EBM.

Campbell and Stanley (1963) posited four aspects of research design that needed to be considered in order to improve internal validity. Firstly, the instrumentation used to measure the variables of interest must be reliable in the measurements obtained across both experimental and control groups. Where the researcher is the research instrument, it was suggested that they should be blind as to which group they were measuring. Secondly, selection bias is addressed through the random allocation of research subjects to the experimental and control groups. Thirdly, what Campbell and Stanley referred to as 'experimental mortality' (in more modern terminology, 'attrition') needed to be addressed by conducting analysis of outcomes on the basis of the groups to which research subjects were originally assigned, even if they dropped out of the study. Finally, extraneous occurrences in the form of events outside of the researcher's control, the passing of time (how can we be sure that effects are due to the experimental intervention and not simply the passage of time?), and reactivity to the intervention need to be considered and the manner in which they are controlled made explicit.

2. External validity:- "The validity of inferences about whether the causal relationship holds over variations in persons, settings, treatment variables, and measurement variables." (Shadish et al., 2002, p.507)

Where internal validity looks inwards, external validity looks outwards – it might be useful to know that a smoking cessation programme results in a significant decrease in the numbers smoking in a middle-class area, but what if your area of concern is working-class smokers, or if you do not have the resources to deliver the programme to the fidelity attained in the study? It is important to know whether the results of the research are sufficiently robust to withstand these differences and the imperfections of day-to-day programme delivery.

Campbell was clear that the problem of external validity could not be solved simply be reference to inferential statistics (Campbell and Russo, 1999). Thus, whilst the problem of selection bias interacting with the experimental variable could jeopardise external validity (Campbell and Stanley, 1963), addressing this alone was insufficient for assuring external validity. In particular, Campbell and Stanley were concerned that the nature of the experimental situation itself might produce outcomes that greatly outweighed that attributable to the intervention *per se*. It was proposed that the key way in which external validity could be improved was through the trial and error testing of the proposed generalisations in other contexts (Campbell and Stanley, 1963; Campbell and Russo, 1999). This was a resource-intensive proposition, but one that Campbell saw no acceptable way around.

3. Construct validity:- "The degree to which inferences are warranted from the observed persons, settings, and cause-and-effect operations sampled within a study to the constructs that these samples represent." (Shadish et al., 2002, p.506)

Although construct validity may appear broadly similar to external validity, it is important to distinguish how it is different. A construct is not a given, it is

something that is devised by a community in order to make sense of phenomena, and constructs about the same phenomenon can legitimately be quite different in different contexts. For example, the distinction between women and men (constructs) is made with regard to researching gender inequalities in earnings, and indeed there are strong reasons for doing so. However, constructs upon which to research differences in earnings could equally be made with regard to individual psychology (perhaps measurements of determination or selfishness), or the extent of the previous generation of the family's involvement in high-paying professions. The important point here is that valid constructs enable us to be explicit about *why* it is possible to make inferences from the particulars of the research sample to certain others; why is the construct of distinguishing men and women considered valid when researching differences in pay? Is it the case that there may be equally valid constructs that would allow us to interpret differences in what people earn along quite different lines? If so, what would be the implications for policy and practice in this area?

It is important to note that research cannot simultaneously fulfil all three validity criteria (Shadish et al., 2002, p.93-102). Stated a little bluntly; research can make very accurate statements about the relationship between tightly circumscribed phenomena, but which cannot be applied in any other context (very high internal validity); it can make very broad statements that remain robust in multiple contexts, but which are possibly so banal as to be of no practical application (very high external validity); or, it can test out many different variants of its constructs in order to explore how this may change our understanding of phenomena, but which says nothing rigorous about the relationship between those phenomena (very high construct validity). Clearly, none of these (extreme) outcomes are desirable, and researchers rarely commit so strongly to one type of validity as to produce such results.

However, what is important is that researchers should explicitly consider the weighting that it is appropriate to give to these different types of validity in their research designs, and explicitly account for their decisions. For example, in testing the effectiveness of a new drug upon a medical condition about which there is considerable understanding, a strong emphasis upon internal validity is likely to be justified - we know very precisely the relationship to be tested and can confidently exclude confounding factors. Where safety is not such an issue (for example, 'talking therapies' with people who are not clinically depressed) and where it is proposed that the intervention be used across a diverse range of contexts, we may legitimately emphasise external validity. Finally, where the suitability of the manner in which we classify the 'objects' of research is questioned, an emphasis upon construct validity may allow the development of new ways of looking at the phenomena concerned which may better inform our understanding of it. For Campbell, a judgement regarding validity as a whole was what really mattered, and should be particularly attuned to the effect that the political pressures inherent to many projects have upon the results obtained (Campbell, 1978).

Campbell furthermore expanded the understanding of the ways in which validity could be attained by firmly locating the construction of

knowledge as something that took place within the social sphere. Campbell firstly meant by this that science depended in large part on the social process of criticism of research by colleagues and the drive that researchers felt to conduct rigorous research in an honest manner (Campbell, 1986). However, Campbell also argued that if science was an inherently social process, then research validity would be improved by better specifying these social processes and the influence they exerted on the observations and judgements that were made in the course of conducting research (Campbell and Russo, 1999).

Conclusion

The key point with regard to the use of internal, external, and construct validity outlined above is that these three types of validity are not simply the preserve of so-called different research 'paradigms'. The debate within EBP has frequently, and very unproductively, polarised along the lines of 'quantitative' (broadly allied with an emphasis upon internal validity) and 'qualitative' (broadly allied with an emphasis upon construct validity)[4], whereas an approach that emphasised selecting the right (mix of) 'tools for the job' would arguably serve researchers, practitioners and policy-makers (and indeed the population as a whole who are subject to the interventions made) much better. The advantage of utilising internal, external and construct validity as 'tools to think with' is that they provide a framework for structuring research that rationally uses different methods, rather than trying to force legitimate research questions into unsuitable methods. Moreover, 'thinking with' validity allows the development of strong rationale for moving backwards and forwards through degrees of emphasis upon different forms of validity in subsequent research projects; the ideal here is that research should increase its rigour still further through becoming part of an ongoing, constructive, debate within a community (Campbell, 1986; Campbell and Russo, 1999) rather than simply the establishment of methodological superiority.

[1] Campbell envisaged such a society as being one that *tested* interventions that were purported to improve people's well-being, rather than relying upon ideology to inform policy-making (Campbell and Russo, 1999).

[2] For example, see the Oxford Centre for Evidence-Based Medicine's 'levels of evidence' (http://www.cebm.net/levels_of_evidence.asp).

[3] By 'requirements', I do not mean that (for example) the predilections of Ministers should dictate the remit of research, nor that (for example) the parochial concerns of practitioners would provide a suitable basis for deciding upon research foci. What is meant is that, if research is to say anything more than 'this is what my research shows in this context', there has to be an explicit justification for why and how the findings are applicable to decision-making outside of the precise contextual conditions within which the research was conducted.

[4] It might be observed here that neither quantitative nor qualitative approaches have been overly concerned with external validity as understood by Campbell and colleagues. Whilst techniques of statistical inference have become highly developed, these inferences do not have a direct (one-to-one) relationship with what occurs in reality; they thus require careful interpretation (Byrne, 2002). Qualitative approaches have often been more concerned with 'breaking free' from the binds of scientific practice than considering how they may be made more rigorous, although clearly the rigour of qualitative research methods has been substantially developed by some (for example, Mason, 2002).

Byrne, D. (2002) *Interpreting Quantitative Data,* London: Sage Publications.

Cabinet Office (1999) *Professional Policy Making for the Twenty First Century,* London: Cabinet Office.

Campbell, D. (1978) 'Qualitative knowing in Action Research', In *The Social Contexts of Method* (Eds, Brenner, P., Marsh, P. and Brenner, M.) London: Croom Helm, 184-209.

Campbell, D. (1986) 'Science's social system of validity-enhancing collective belief change and the problems of the social sciences', In *Metatheory in Social Science* (Eds, Fiske, D. W. and Shweder, R. A.) Chicago: University of Chicago Press, 108-135.

Campbell, D. and Russo, M. J. (1999) *Social Experimentation,* London: Sage Publications.

Campbell, D. and Stanley, J. (1963) *Experimental and Quasi-Experimental Designs for Research,* Boston: Houghton Mifflin Company.

Chalmers, A. and Altman, D. (Eds.) (1995) *Systematic Reviews,* London: BMJ Publishing Group.

Chalmers, I. (2003) 'Trying to do more good than harm in policy and practice: the role of rigorous, transparent, up-to-date evaluations', *Annals of the American Academy of Political and Social Science,* 589 (1), 22-40.

Chalmers, I. (2005) 'If evidence-informed policy works in practice, does it matter if it doesn't work in theory?' *Evidence & Policy,* 1 (2), 227-42.

Clegg, S. (2005) 'Evidence-based practice in educational research: a critical realist critique of systematic review', *British Journal of Sociology of Education,* 26 (3), 415-28.

Cook, T. D. and Campbell, D. (1979) *Quasi-Experimentation,* Chicago: Rand McNally.

Department of Health (1999) *Saving Lives: Our Healthier Nation. Cm4386,* London: The Stationery Office.

Gabe, J., Kelleher, D. and Williams, G. (Eds.) (1994) *Challenging Medicine,* London: Routledge.

Glasziou, P. and Haynes, B. (2005) 'The paths from research to improved health outcomes', *Evidence-Based Medicine,* 10 (1), 4-7.

Goodman, N. W. (1999) 'Who will challenge evidence-based medicine?' *Journal of the Royal College of Physicians,* 33 (3), 249-51.

Hammersley, M. (2004) 'Some questions about evidence-based practice in education', In *Evidence-Based Practice in Education* (Eds, Thomas, G. and Pring, R.) Maidenhead: Open University Press, 133-149.

Hammersley, M. (2005) 'Is the evidence-based practice movement doing more good than harm? Reflections on Iain Chalmers' case for research-based policy making and practice', *Evidence & Policy,* 1 (1), 85-100.

Haynes, B. and Haines, A. (1998) 'Getting research findings into practice: barriers and bridges to evidence-based clinical practice', *British Medical Journal,* 317 273-6.

Holmes, D., Murray, S. J., Peron, A. and Rail, G. (2006) 'Deconstructing the evidence-based discourse in health sciences: truth, power and fascism', *International Journal of Evidence Based Healthcare,* 4 (3), 180-186.

Home Office (1999) *Reducing Crime and Tackling its Causes: A Briefing Note on the Crime Reduction Programme,* London: Home Office.

Kelly, M. (2006) 'Mapping the life-world: a future research priority for public health', In *Public Health Evidence: Tackling Health Inequalities* (Eds, Killoran, A., Swann, C. and Kelly, M.) Oxford: Oxford University Press, 553-573.

Mason, J. (2002) *Qualitative Researching,* London: Sage Publications.

Oakley, A. (2006) 'Resistances to 'new' technologies of evaluation: Education research in the UK as a case study', *Evidence & Policy,* 2 (1), 63-87.

Oakley, A., Gough, D., Oliver, S. and Thomas, J. (2005) 'The politics of evidence and methodology: lessons from the EPPI-centre', *Evidence & Policy,* 1 (1), 5-31.

Pietroni, P., Winkler, F. and Graham, L. (2003) 'Cultural revolution', *British Medical Journal,* 326 (7404), 1304-6.

Polychronis, A., Miles, A. and Bentley, P. (1996) 'The protagonists of 'evidence-based medicine': arrogant, seductive and controversial', *Journal of Evaluation in Clinical Practice,* 2 (1), 9-12.

Shadish, W. R., Cook, T. D. and Campbell, D. (2002) *Experimental and Quasi-Experimental Designs for Generalized Causal Inference,* Boston: Houghton Mifflin Company.

Tannenbaum, S. J. (1993) 'What physicians know', *New England Journal of Medicine,* 329 (17), 1268-1271.

Tilley, N. (2001) 'Evaluation and evidence-led crime reduction policy and practice', In *Crime, Disorder and Community Safety: A New Agenda?* (Eds, Matthews, R. and Pitts, J.) London: Routledge, 81-97.

Victora, C. G., Habicht, J.-P. and Bryce, J. (2004) 'Evidence-Based Public Health: moving beyond randomized trials', *American Journal of Public Health,* 94 (3), 400-8.

Williams, S. J. and Calnan, M. (1996) 'The 'limits' of medicalization?: Modern medicine and the lay populace in 'late' modernity', *Social Science & Medicine,* 42 (12), 1609-20.

©The Policy Press • 2007 • ISSN 1744 2648

Systematic reviews in social policy: to go forward, do we first need to look back?

Mark Pearson

English This article argues that the development of systematic review methods for social policy and practice in the UK has been inhibited in two ways. First, there has been insufficient engagement with the breadth of Donald Campbell's thought on research methods and knowledge cumulation. To this end, the article outlines the contribution that a fuller understanding of Donald Campbell's work could make to the development of such methods. Second, debates about the merits of systematic review methods have often shown only a partial understanding of other approaches. The article thus seeks to identify key areas of commonality and difference between approaches to systematic reviews with a view to stimulating methodological development through constructive debate.

Français Cet article défend le point de vue que le développement des méthodes de revues systématiques en politique et en pratique sociale au Royaume Uni a été entravé de deux façons. D'abord, on n'a pas suffisamment largement tenu compte de la pensée de Donald Campbell sur les méthodes de recherche et le cumul de connaissances. Dans ce but, l'article donne les grandes lignes de la contribution qu'une meilleure compréhension du travail de Donald Campbell pourrait faire au développement de telles méthodes. Deuxièmement, les débats sur les mérites des méthodes de revues systématiques n'ont souvent montré qu'une compréhension partielle d'autres approches. L'article cherche ainsi à identifier les domaines clé de similitude et de différence entre les approches aux revues systématiques dans le but de stimuler un développement méthodologique à travers un débat constructif.

Español Este artículo discute que el desarrollo de métodos de revisión sistemática para la política social y práctica en el Reino Unido ha sido inhibida de dos maneras. Primero, no ha habido suficiente compromiso con la amplitud del pensamiento de Donald Campbell en métodos de investigación y acumulación de conocimiento. Con este fin, el artículo destaca que la contribución que un mayor entendimiento del trabajo de Donald Campbell podría hacer que se desarrollasen tales métodos. Segundo, los debates acerca de los méritos de los métodos de revisión sistemática han mostrado con frecuencia sólo un entendimiento parcial de otros enfoques. El artículo por lo tanto busca identificar áreas claves de similitud y diferencia entre enfoques de revisiones sistemáticas con una opinión para estimular desarrollo metodológico a través de un debate constructivo.

Key words systemic reviews • evidence-based policy and practice • Donald Campbell

Comprehensively collating and synthesising knowledge on a topic by means of a systematic review (SR) is the bedrock of evidence-based policy and practice (EBP), and moreover is an endeavour against which no serious case has been made. Grounded in pioneering work that challenged the perpetuation of harmful practices in clinical medicine (Evidence-Based Medicine Working Group, 1992; Sackett et al, 1996; Clarke and Chalmers, 1998), the concept of systematically reviewing and cumulating evidence resonated widely within a UK government keen to consolidate its pragmatic credentials (Cabinet Office, 1999; DH, 1999; Home Office, 1999). However, despite this consensus over the desirability of EBP, debate about what constitutes rigorous knowledge and how it may best be cumulated has at times been problematic (for example, Chalmers, 2005; Hammersley, 2005).

Latterly, key figures in the field of EBP in the UK have made substantive efforts to further the integration of different forms of evidence and to re-evaluate the appropriateness of the hierarchy of evidence developed under the rubric of evidence-based medicine (EBM)[1] (Oliver et al, 2005; Boaz et al, 2006; Littlejohns and Chalkidou, 2006). While such efforts are to be applauded, they arguably provide more of a springboard for debate than an in-depth consideration of a number of substantive underlying issues. This article endeavours to take the next step by offering an analysis of these underlying issues by expanding on the areas of agreement and disagreement between different approaches to conducting SRs in social policy. It is intended that this analysis should enable substantive debate on the issues, and facilitate methodological development. Without the development of this kind of understanding, SR teams are likely to maintain their 'methodological preferences' because they are comfortable with them, rather than because they are necessarily the best tools for the job (Boaz et al, 2006).

As a basis for considering the different approaches to conducting SRs, this article takes a step further back in order to analyse the work of Donald Campbell. Despite Campbell's name being quoted approvingly in much of the literature on conducting SRs, there has been only limited engagement with the breadth of his thought on issues surrounding research on the effectiveness of interventions in the social arena and on the social nature of all scientific endeavour. Given the difficulties encountered in conducting SRs in the field of social policy and practice, a genuine engagement with Campbell and colleagues' work is of great importance if these issues are to be adequately addressed.

The work of Donald Campbell – lessons for contemporary systematic reviews in social policy?

Donald Campbell's work on evaluative research designs in the field of social policy, and wider issues of scientific practice, risks being all things to all people if only looked at in part. For those who favour randomised controlled trials (RCTs), much will be found in Campbell and Stanley (1963) to bolster the case for this approach producing the most dependable form of knowledge. Those who view the social world as too complex for true experimental designs to be implemented will also find much to support their case in Cook and Campbell (1979) regarding rigorous quasi-experimental designs.

Furthermore, qualitative researchers may be keen to use Campbell's (1978) thinking on the qualitative basis on which all quantitative knowledge is built when debating the relative certainties of different forms of knowledge. Even postmodernists may find something in Campbell's work to support their case in his guarded support for constructivists' work on the social conduct of science (Campbell and Russo, 1999).

The above broad summary risks giving the impression that Campbell's work lacked coherence, that he latched onto trends in the research field rather than ploughing a consistent furrow. Another possible explanation is that Campbell 'moved on', that he abandoned his earlier experimental approach in favour of one that leant heavily towards a social constructivist position. However, these interpretations are not tenable. What marks out Campbell's work is arguably his willingness to traverse supposedly incommensurable approaches in a genuine effort to foster the production of pragmatic knowledge. His advocacy of the experimental approach never diminished as such, it simply became a little more refined and tempered with the knowledge that uncertainties in evaluations arose not only from the details of statistical inference, but also from the social processes inherent in the conduct of scientific research. Campbell's goal as a methodologist was to define a 'course between the extremes of inert scepticism and naïve credulity' (Campbell, 1978, p185), and it is in this spirit that the entire body of his work should be approached, neither discounting Campbell's pursuit of rigour as unattainable nor accepting his advocacy of particular methods as the final word on the subject.

This section will now endeavour to engage with the breadth of Campbell and colleagues' work in order to assess how it may contribute to the contemporary conduct of SRs in social policy. It will be argued that a significant number of the current conflicts over the best methods for performing SRs in social policy have been substantively addressed by Campbell and colleagues. This is not to suggest that a straightforward solution sits waiting within this body of work, but engagement with it would facilitate the contemporary development of SR methods, and would represent a more comprehensive treatment of the literature relevant to the rigorous conduct of SRs in the field of social policy.

Validity – internal, external and construct

Throughout his work, Campbell strongly advocated the experimental approach (in the form of RCTs where possible) to evaluating the effectiveness of social policy interventions because of its strength in discerning causal relationships between phenomena (Campbell and Russo, 1999). For example, the majority of the text of Campbell and Stanley (1963) is devoted to describing different experimental research designs and their various capabilities in addressing threats to internal validity. The importance attributed to the experimental approach remains in Campbell (1978), even though the bulk of that text is about 'qualitative knowing'. Similarly, in Cook and Campbell (1979) the focus on quasi-experimental research designs should not be seen as a dismissal of 'true' experimental designs, so much as a concerted effort to improve the practice of quasi-experimental methodologies. Campbell's advocacy of experimental approaches should thus be understood both in the context of his grasp of the appropriate utilisation of other methodologies and of his increasing consideration

of the importance of external validity, and the inherent tension that exists between assuring high internal validity and high external validity (Shadish et al, 2002).

In his later work, Campbell's advocacy of the experimental approach was a little more guarded, and he conceded that a rigorous utilisation of the method is best suited to tightly controlled environments, such as within prisons (Campbell and Russo, 1999). He also identified a tension between the assignment of people to experimental and control groups and his vision for the 'Experimenting Society', the philosophy of which was egalitarian and voluntaristic. Despite these reservations, Campbell continued to advocate the judicious use of the experimental approach, where it did not conflict with these ideals, for the evaluation of interventions in the social sphere (Campbell and Russo, 1999, pp 24-6). In a similar spirit of egalitarianism, he also opened up the possibility of the involvement of non-researchers in debating research findings, pointing out that those with first-hand experience of phenomena are often in the strongest position to critique the rigour of a piece of research and the causal mechanisms that it purports to demonstrate (Campbell and Russo, 1999, p 37).

Campbell arguably attributed an increasing amount of importance to external validity in his later work, but it would be doing him a disservice to claim that his earlier work did not consider it at all. The difficulties of generalising research findings to other populations or settings are touched upon (but never really resolved) in Campbell and Stanley (1963), where generalisations are seen to inherently involve assumptions about the mechanisms by which causal relationships operate and thus the extent to which these relationships may persist in other contexts. Campbell and Stanley (1963, pp 17-19, 33-4) proposed that through the scientific process, these mechanisms may be further tested, and theories about them revised or refined, but that there was no straightforward 'solution' to the degree of assumption in making generalisations. Fundamentally, what was necessary was a further testing out of theories about causal mechanisms in a range of contexts, and with the knowledge that the social world continued to change and develop apace.

Although recognising the difficulties of making generalisations from research findings to a complex, dynamic social world, Campbell was certain that doing so was a vital task if a cumulative approach to knowledge was to be taken. While he was strongly aware that the social sciences were unlikely to be able to develop laws that were robust over time and place (as in the physical sciences), he argued that it was imperative that the attempt should be made to develop research-based theories capable, to a certain extent, of generalisation (Campbell and Russo, 1999, p 195). Campbell was also clear that statistical inference, in and of itself, was insufficient as a basis for generalisation, particularly where changes in phenomena over time produced as much substantive difference in results as differences attributable to other factors (Campbell and Russo, 1999, pp 116-17). Instead, he proposed purposive sampling based on explicit conceptual criteria that endeavoured to test generalisability claims (Campbell and Russo, 1999, p 121).

A key development in Campbell and colleagues' thoughts on rigorous generalisation was the concept of construct validity, as something distinct from external validity (Shadish et al, 2002). While external validity is concerned with the robustness of causal relationships in different contexts, construct validity focuses on the inferences made

from samples to the 'higher-order constructs' utilised in a field of study . For example, while it may be useful to know that the direction of the relationship between the provision of smoking cessation services and the number of people stopping smoking is robust over different segments of the population (external validity), it is the validity of the constructs about the phenomena concerned (such as people classified by social class) that allow or inhibit rigorous generalisation to other instances. Constructs are not absolutes, they are (or ideally, should be) the subject of debate and development within a field that constantly seeks to re-evaluate the appropriateness of its conceptualisations (Shadish et al, 2002, pp 66-72).

The transparency of research methods

Campbell's methodological proposals always sought to attain as much transparency as possible, but he was also quite explicit about the substantive tacit element, both of his own methods and those of others. This tacit element was first outlined by Campbell in his work on the qualitative basis of all scientific knowledge, which he suggested went mostly unrecognised simply because of its ubiquity (Campbell, 1978). One of Campbell's concerns was that a spurious precision could be apparent in research that was based on unexamined qualitative knowledge, for example where the coding of interview responses is subsequently quantified for analysis. The impact that this qualitative knowledge (potentially a mixture of 'common-sense', professional training, and personal experience and reflection) has on the results of a study (wherever it may lie on, for example, an experimental–ethnographic continuum) is rarely considered, yet it may be of substantive import (Shadish et al, 2002, p 29).

Campbell argued strongly that these underlying assumptions should be investigated in order to make experimental research more genuinely scientific, his proposal for a 'project anthropologist' (Campbell, 1978) being realised in the work of social constructivist researchers who examined the social process of science within laboratories (Knorr-Cetina, 1981; Latour and Woolgar, 1986). Perhaps surprisingly, he wrote approvingly of the work of these researchers (Campbell, 1986, pp 112-19) and the manner in which they brought forth the social processes by which laboratory researchers imposed order 'on a chaotic welter of inconsistent and inconclusive observations' (Campbell, 1986, p 118). Campbell was clear that such social processes did not invalidate the knowledge produced, but that overlooking the role that these social processes played *did* reduce scientific validity, as their role was not made clear (Campbell, 1986).

The role of the research community

Campbell argued that the conduct of rigorous science relied more on a process of critical monitoring by the research community than the following of explicit procedures (Campbell and Russo, 1999). He referred to this as 'competitive cross-validation', a social process that requires scientific claims to validity to be justified to a sceptical community, rather than taken on trust on the basis of a belief in a particular researcher's honesty or competence (Campbell, 1986).

The critical attentiveness that Campbell viewed as so vital for improving the validity of scientific claims worked through three different mechanisms (Campbell and Russo, 1999). First, in a community in which members respect and value each other's contributions, Campbell posited that the rigour of scientific conduct is improved by the fear of humiliation if others' attempts to replicate an experiment fail because of weaknesses in the original design. Second, the nature of the community motivates researchers to compete with one another in the sense of ensuring that high standards of conduct are maintained. Third, the critical nature of the community is such that it relies for its reputation on maintaining rigorous standards, and thus is more likely to disclose poor practice than to cover it up. A key factor in all of these mechanisms is the focus on the community, the 'disputatious community of "truth seekers"' as Campbell coined it (Campbell and Russo, 1999, p 9), in entering into debate. It was not a matter of procedure assuring knowledge of higher validity; rather it was a contingent process of informed debate that necessarily considered both internal and external validity and, through mutual persuasion, arrived at an agreed conclusion (Campbell, 1986; Campbell and Russo, 1999).

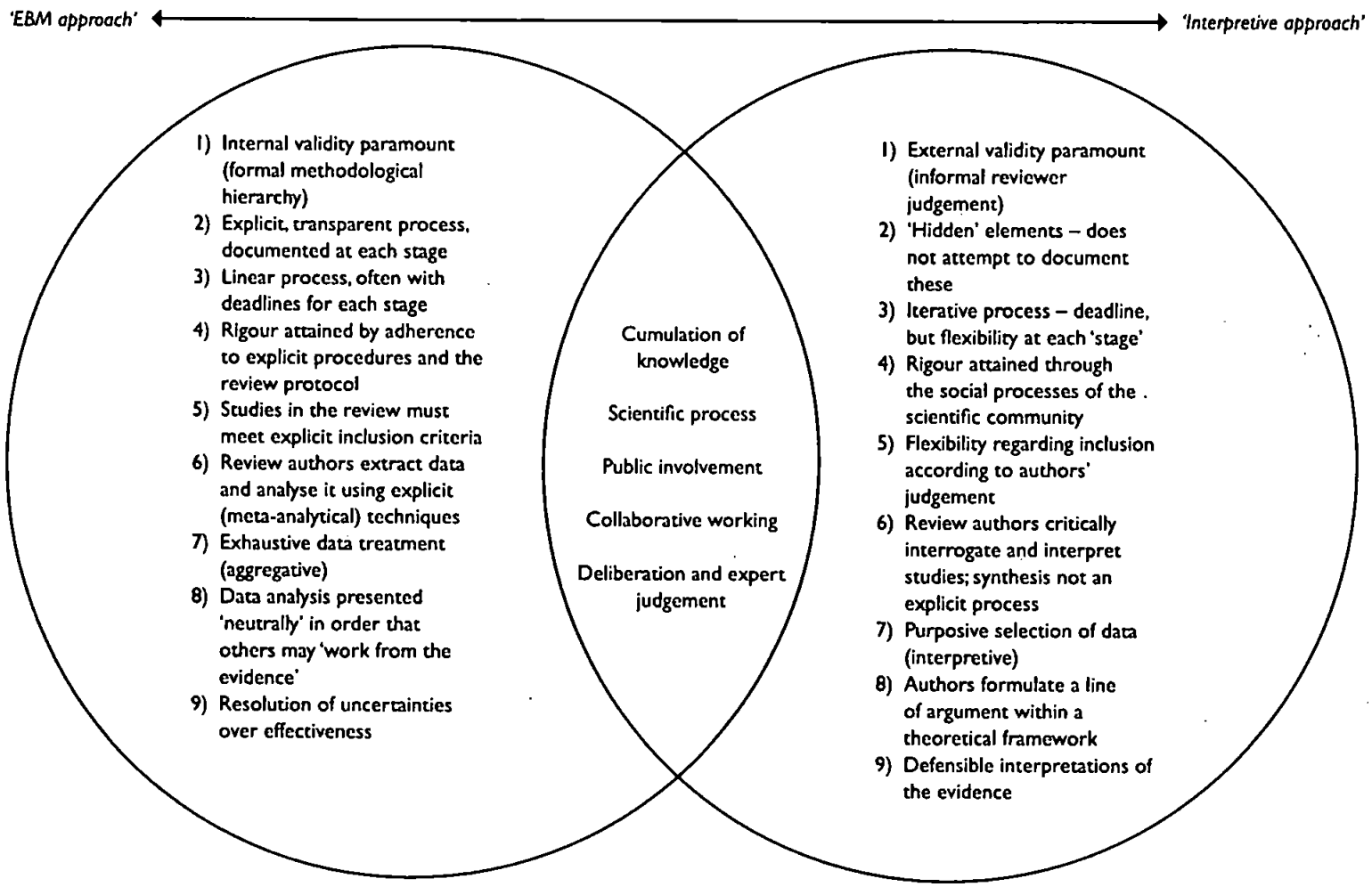
The importance of pattern identification

Campbell was clear that it was in the nature of knowledge that it cannot be lifted unproblematically out of the context in which it was produced: necessarily, the interpretation of data demands a qualitative knowing both of this context and of the manner in which the data can be made sense of within the field of knowledge as a whole ('pattern identification', see Campbell [1978, p 191]). While Campbell strongly argued in his conception of the 'Experimenting Society' that one should be willing and able to change one's views in the light of evidence (Campbell and Russo, 1999), he did not propose that the way to do this was to start from isolated data ('the very reverse of dependable building blocks'; Campbell, 1978, p 191). Rather, an informed and careful interpretation of these data, in which the way pattern identification is utilised is made as explicit as possible, is contended by Campbell to be a superior (if more difficult) way to 'work from the evidence'.

Approaches to social policy SRs in the UK

This section provides an overview of the current range of approaches to conducting SRs.[2] It is intended to be read with Campbell's consideration of the validity and transparency of research methods, and of the role of the research community, being borne in mind. First, the broad areas of agreement about objectives in the field, which are arguably greater than is generally acknowledged, are identified. By expanding on these commonalities, I hope to provide a firm basis for critically examining the often polarised methods proposed in the different approaches to SRs. This critical examination is intended to highlight the underlying rationale behind the different methods in order to better understand *why* these polarised approaches are strongly advocated. The structure of this analysis, together with a visual representation of where some key approaches lie on the spectrum, is presented diagrammatically in Figure 1.

Figure 1: The spectrum of approaches to conducting a systematic review



Areas of agreement

The cumulation of knowledge

A key aspect of the rationale for SRs is that policy and practice can be much better informed if research on a particular topic is analysed comprehensively, rather than in a piecemeal and potentially partisan way (Chalmers and Altman, 1995; Sackett et al, 1997; Egger et al, 2001a). The social sciences have been criticised strongly for not developing an institutional structure that is supportive of such an approach, and its lack is said to result in the needless repetition of research, or the basing of research on knowledge that would have been discredited if a comprehensive SR had been undertaken (Chalmers, 2003; Oakley et al, 2005; Petticrew and Roberts, 2006). Moreover, this lack of institutional development is viewed by some as the result of active resistance to methods that threaten established ways of working and vested interests (Oakley, 2006).

From this viewpoint, there does not seem to be much common ground between perspectives: the social sciences are seen to have 'lost their way' with the advent of postmodernism, and stand accused of effectively throwing away a strong tradition of a genuinely scientific, cumulative approach to knowledge (Oakley et al, 2005). While mindful of some of the more extreme postmodernist approaches, the more moderate social scientists that are the subjects of Oakley et al's critique do not appear, on further examination, to be so different with regard to the cumulation of knowledge. Writing in 1985, Martyn Hammersley voiced his concern that research in the sociology of education was proceeding along disorganised lines, with researchers failing to build on earlier work, synthesise their own research findings with those of others or work collaboratively (Hammersley, 1985).

Arguably, the differences between Oakley and Hammersley relate to the way in which they view knowledge cumulating. Oakley's position is shared with that of Iain Chalmers, where the effect sizes computed literally cumulate: the more data one has, the narrower the confidence interval and thus the greater certitude with which statistical inferences can be made. Hammersley (2001) does not dispute this view of knowledge cumulation, but proposes that it is not the only way in which knowledge may cumulate – it can also be additive in the sense of forming a 'mosaic', or it can challenge or re-enforce other knowledge (Hammersley, 2001, p 548). Hammersley's position is thus shared with that of Freese (1980), who argued that statements about relationships between phenomena do not constitute a science, unless it can be demonstrated how these relationships operate. Such an understanding requires more of cumulation than is possible with a purely additive approach, but the difficulties it poses for maintaining transparency and accuracy can be serious (Sharland and Taylor, 2006). Nevertheless, there is congruence between the approaches in that they both view it as essential that the production of knowledge takes place within an organised framework that prevents repetition and facilitates the utilisation of prior research.

The scientific process

It could be argued that debates over SR methodologies are incommensurably rooted in different research paradigms, but there may be more common ground between the approaches than might be evident on first examination. This common ground exists in three areas: the underlying tenets of 'good' science; how theories should be tested; and the role of judgement in the scientific process.

First, there is common ground on how science should proceed: evidence should be provided for the conclusions that are reached, and methods should be reported so that the research community can critique the work reported on (Oakley, 2000; Hammersley, 2001). Second, despite the apparent gulf between the experimental approach to research and, for example, ethnographic research, there is arguably more common ground than is normally acknowledged. Writing in 2003, Iain Chalmers summarised the rationale behind his longstanding advocacy of the experimental approach in both medical and social policy research. In doing so, he expanded on the scientific approach whereby theory is refined or rejected on the basis of research findings, and where the essence of the scientific approach with respect to establishing causation is the manipulation of variables so that relationships between phenomena may be observed (Chalmers, 2003, p 27). Almost 20 years previously, long before the fields of the two authors had come into contact with each other, Martyn Hammersley was writing of the importance in educational research of the 'testing out of predictions derived from [a] theoretical idea' (Hammersely, 1985, p 247). While Chalmers and Hammersley remain divided over the best means of establishing causation (for Chalmers, RCTs; for Hammersley, ethnographic research that tests out theory), they are not at odds over the essential activity of manipulating phenomena so that the (conditional) effects can be observed or interpreted.

Finally, there exists a nominal conflict over the role of judgement in the conduct of research, and in particular the conduct of SRs. Hammersley characterises the 'positivist' approach as basing validity on adherence to explicit procedures and rules, thus factoring out what he views as the inherent (and important) role of human judgement in the scientific process (Hammersley, 2001, 2005). However, Oakley (2000, p 302) asserts that such a characterisation is unwarranted because it fails to appreciate the 'messy, non-linear, creative [and] contingent' nature of the conduct of all science. While these two authors may still disagree over the extent to which judgement is actually exercised in scientific practice, there is no dispute about the role that human judgement, rather than uncritical rule-following, may play in the scientific process.

Widening involvement in the SR process

Intricately tied up with the advocacy of SRs is the idea of democratising knowledge. In a world where access to information is hugely facilitated by computer technology and telecommunications, and where there is an increasing public demand for the tools that allow devolved decision making (as opposed to entrusting decisions to expert bodies), SRs have been promoted as a key method for synthesising research evidence

and making it available publicly. There is arguably broad agreement within the field regarding the involvement of the public at various key stages of a SR, for example to refine the initial review question, to provide feedback on progress at interim stages of the review and to comment on the practical relevance of the final review's findings. While the substance of efforts made to encourage and facilitate service users and the public to contribute to a review as equal partners (for example, see Braye and Preston-Shoot, 2005) are unlikely to be contested, a sociologically rich understanding of the methods of consultation, how they affect the views expressed and how these are integrated with expert knowledge is arguably missing. For example, Harden (2001) reflected on some of the difficulties of eliciting and balancing views that were expressed in a review's consultation with young people, but further work on this problematic area has not been forthcoming. Nevertheless, the basic commitment to public involvement in the conduct of SRs is shared across the field.

Despite this shared commitment, it should be acknowledged that differences do exist with regard to accessing and utilising the knowledge so produced. For example, Oakley et al (2005) view the collation of SRs in publicly accessible electronic libraries as essential for the democratisation of knowledge: not to do so simply protects the power of professionals and places the service user in a subservient position, unable to question the professional's 'expertise'. A different view is held by others: for example, Kuhlmann (2004) maintains that the democratisation of knowledge needs to take place at a more fundamental level, when what is acceptable as evidence or knowledge is decided upon. While this substantive difference in the field persists, there is at least the potential for development in that both perspectives highly value the role of those outside of the 'expert community'.

Collaborative working

Criticisms regarding the lack of individual and institutional response to the call for conducting SRs have focused strongly on a perceived lack of understanding of the rationale for doing so. Chalmers et al (2002) strongly criticise academic practice in the UK for not seriously considering the imperatives of research synthesis and how the challenges it presents may best be responded to. Oakley et al (2005) saw little development in academics' approach in the intervening years, criticising academia in the UK for hindering the cumulation of knowledge by conducting research projects that were 'parochial' and which lacked 'any sense of connectedness with one another' (Oakley et al, 2005, p 21). It is debatable whether these critiques apply as strongly now as they did at the time they were made; for example, the review approaches of Dixon-Woods et al (2006) and Pawson (2006) clearly highlight the importance of situating a review in the context of other work in order that a piecemeal approach is not adopted. Pawson (2006) also goes a step further by specifying an approach in which collaborative working between reviewers, commissioners and policy makers enables the review to be kept highly relevant to the needs of these parties. While the ideal of a wholly collaborative social policy review community has certainly not been attained, there is again the common ground that acknowledges the importance of collaboration on which to build.

Deliberation and expert judgement

Systematic reviews, particularly in healthcare, have at times been unfairly portrayed as proceeding along purely rule-bound tracks that prohibit certain forms of knowledge from being included and exercised, such as that inherent in professional judgement (for an extreme example, see Holmes et al, 2006). However, the classical formulation of EBM explicitly acknowledged that decision making should integrate more tacit expertise with the explicit research synthesis conducted in a SR (Sackett et al, 1996). Although the actual conduct of EBM *might* have emphasised the latter, key figureheads continue to highlight the importance of expert clinical judgement (for example, Haynes, 2005) and of political judgement that is mindful of the role of policy makers as representatives of the populace (for example, Gray, 2005) in decision making that is informed by evidence.

It is arguably now widely recognised that EBP has to make substantial use of both the syntheses of SRs *and* professional judgement in the deliberations that inform decision making. However, a cautionary note should be sounded, for although there may be a convergence in views about the importance of judgement in EBP, the stages at which it is considered to be appropriately exercised differ quite markedly. For example, Chalmers' (2003) starting point is that judgements will have to be made that are mindful of resources and values, as well as 'the evidence' (Chalmers, 2003, p 36), but judgements do not enter into the picture until the point at which 'the evidence' has been established. This sharply contrasts with Hammersley (2005), who contends that judgement is exercised constantly throughout the review process as well as in the primary research that is being synthesised. Hammersley brings the role of judgement into the foreground in order to highlight the manner in which exercising judgement is inherently problematic, but he also asserts that this does not invalidate it in the way that Chalmers asserts. Thus, there are foundations on which to build regarding the exercising of judgement in SRs, but also a number of substantive issues that require debate.

The above analysis has endeavoured to identify areas of potential agreement in an effort to provide a framework that will allow advocates of different approaches to SRs to see that their respective goals (if not necessarily their proposed methods) have more in common than is usually acknowledged. Where the approaches differ, I have endeavoured to identify exactly where the disagreements are in a bid to foster substantive debate. The following section does not attempt to reconcile different approaches to conducting SRs but to make clearer the rationale for *why* different methods are advocated, on the basis that a better mutual understanding will allow advocates of different approaches to engage with each other more constructively.

Areas of disagreement

Internal versus external validity

> Internal validity: 'The validity of inferences about whether the relationship between two variables is causal' (Shadish et al, 2002, p 508).
>
> External validity: 'The validity of inferences about whether the causal relationship holds over variations in persons, settings, treatment variables, and measurement variables' (Shadish et al, 2002, p 507).

In view of the multiple sources of potential bias in the conduct of primary research, internal validity is argued to be the primary consideration in judging what research should be included in a SR (Chalmers et al, 2002). It is argued that bias will impinge on the process of research at a number of stages if steps are not taken to address it. Ensuring that research subjects are unaware of whether they are receiving an experimental or standard treatment (single-blinding), or that both research subjects and investigators are unaware (double-blinding), is proposed to prevent performance bias (where subjects alter their behaviour in response to the treatment they are receiving) and detection bias (where investigators' assessments of results are biased towards their own views of what the results 'should' be) (Davies and Nutley, 2000; Egger et al, 2001b). Systematic differences between groups of research subjects may result in certain subgroups being far more likely to drop out of a study than others and this attrition needs to be factored into the analysis if it is not to be skewed (Egger et al, 2001b). Finally, selection bias is overcome by the initial random allocation of research subjects to intervention and control groups. Randomisation is contended to not only prevent investigators assigning subjects to a particular group for unstated reasons (perhaps they may feel that a subject will benefit greatly from an intervention, and thus contribute to demonstrating an intervention's effectiveness), but also to produce experimental and control groups that are equivalent, with (assuming an adequate sample size) both known and unknown confounders distributed equally between the two groups (Oakley, 2000; Higgins and Green, 2005). It may also be argued that the greater the complexity of the phenomena of concern, the greater the need for random allocation to experimental and control groups in order to distribute these confounding factors equally.

The prioritisation of internal validity has, however, been critiqued by those who are otherwise strong advocates of the role of randomisation in attaining it. Glasziou et al (2004), while maintaining that randomisation is the strongest method by which to attain internal validity, also clearly state that the ranking of studies on the basis of internal validity may lead to important qualities in other research designs being overlooked. These qualities may include a deeper understanding of people's motivations and rationale (qualitative research), the ability to evaluate an overall outcome over a long time period (cohort studies) and the identification of novel positive or negative effects that warrant further investigation (case reports).

More fundamental critiques of the utility of randomisation have also been made regarding the logic behind the claim that randomisation (if properly conducted) overcomes bias. First, the claim that evenly distributing confounders (both known and unknown) are adequately attained through randomisation is contested by Alan Chalmers (1999). He suggests that the essence of the experimental approach is that every effort will be made to identify potential confounders, and an explicit account given of how those confounders are removed or controlled for. In this view, relying on randomisation to distribute these confounders is an abrogation of the researcher's responsibility. Second, where research on complex phenomena is conducted, there may be important confounding factors that are *unknown*. If an indefinite number of these unknown confounders exists, how is a researcher to assess the probability of the experimental and control groups being unbalanced through the play of chance (Worrall, 2002, pp 9-12)?[3]

There have also been criticisms of the use of internal validity as the primary criterion for objectively assessing the rigour of research when conducting a SR. This critique considers the objectivity of this process to be illusory in view of the knowledge about the subject area that will be brought to bear on any assessment of research validity (Hammersley, 2001, 2004). In this view, objectivity is not attained through adherence to procedures for assessing internal validity because to do so leads to an impoverished assessment of the evidence. Utilising judgement in the process is argued to enable a more informed appreciation of validity, but it is achieved by a process that is not explicit in the sense in which the EBM approach requires.

Advocates of prioritising the external validity of primary research focus on the practicalities of everyday delivery of public services and the impact of contextual factors on outcomes. Writing on the field of special education, Gallagher (2004) questions how a study using the experimental approach can establish that implementation fidelity was attained: how do researchers purport to have separated out the delivery of an intervention and, for example, the interpersonal skills of a teacher with a group of children? In asking this, Gallagher is making a case for the importance of external validity, and arguably for some less precise evidence that is more amenable to judicious application in other classrooms. Victora et al (2004) make a similar case in respect of evidence for public health interventions, arguing that evaluations of interventions in a variety of everyday settings accompanied by a plausible rationale for why the outcomes observed occurred, would better inform policy and practice than studies with the highest internal validity. Bambra (2005) highlights the difficulties inherent in attributing impact to specific policies (and thus their potential impact in other contexts) when the environment in which these policies are evaluated is already shaped by other policies that are themselves evolving. In a similar vein, Wallace et al (2006) point to the significant social and economic changes that can take place over time and which may confound external validity.

The importance of considering the effect of contextual factors on outcomes is stressed by those who make the case for greater weight to be given to external validity. Hammersley (2005) makes the general point that in utilising the findings of research, due consideration should be given to the representativeness of the sample in relation to the wider population to which the findings are proposed to be applied.

Such considerations of context are arguably rooted in work on realist evaluation, which posed the question, 'What works for whom in what circumstances?' (Pawson and Tilley, 1997; Pawson, 2002). This is especially important in the field of public health, where the focus on internal validity has resulted in a lack of consideration of how key drivers of health inequalities (such as gender, ethnicity or place) affect people's responses to interventions designed to promote health (Killoran and Kelly, 2004; Weightman et al, 2005; Asthana and Halliday, 2006).

Contemporary developments in methods for conducting SRs have pushed the role of external validity up the agenda. Oliver et al (2005) advocate a 'weight of evidence' approach where both internal and external validity are assessed for their adequacy in answering the review question, and a judgement is made about how much weight should be given to the findings of particular studies in formulating the synthesis of evidence. Bonell et al (2006) propose that RCTs should routinely include process evaluations that highlight the degree to which findings may be generalised to a wider population, and which investigate how contextual factors may affect outcomes and the mechanisms through which these effects occur. While these approaches represent some convergence in thinking on issues of internal and external validity, they arguably avoid consideration of more fundamental issues such as the appropriateness of the hierarchy of evidence and the precise nature of evidence synthesis.

More fundamental critiques and radical proposals for methodological development come from Dixon-Woods et al (2006) and Pawson (2006). Both of these proposals endeavour to balance a (non-formalised) judgement of internal validity with consideration of how useful a particular piece of research is likely to be in answering the review question, and differ from conventional narrative reviews with respect to the efforts made to make the procedure transparent. These methods do not attempt the (logically implausible) use of research with both the highest internal validity *and* external validity;[4] rather, they allow for the judicious use of different parts of different studies by the reviewer(s) in order to develop concepts (Dixon-Woods et al, 2006) or middle-range theories (Pawson, 2006).

Sequential versus iterative review procedures

A central concern of SRs is that they should bring a rigorous scientific process to bear on the task of synthesising evidence, and so eliminate the biases that have resulted in effective interventions remaining unrecognised, and ineffective (and possibly dangerous) interventions continuing unchecked (Chalmers and Altman, 1995; Egger et al, 2001a; Higgins and Green, 2005). The EBM approach to SRs proposes that the appropriate way to do this is to have an explicit procedure formalised before commencing a review; this procedure should state the review's objectives and exactly how evidence will be sought, on what basis it will be included, and how it will be synthesised and analysed (Chalmers, 2003). Decision making during the review is made explicit through the utilisation of a data extraction form, which records not only data but also disagreements between reviewers and how these were resolved (Higgins and Green, 2005). This record may be utilised by others to inform a re-analysis of the data if the rigour of the review comes into question. These procedures

are designed to best address the deleterious effects of bias and, in conjunction with the play of chance being addressed by randomisation, produce knowledge in which one may have more faith (Chalmers, 2003).

The transparency attained by adhering to an explicit, sequential process for conducting SRs has, however, been questioned on a number of fronts. First, with regard to meta-analyses, the sheer volume of data involved is argued to produce a great risk of significant errors occurring, which the critical reader has scant possibility of investigating and questioning (Slavin, 1995; Briggs, 2005). The volume of data may also drown out the role that the reviewers' judgement played, again inhibiting critique of the meta-analysis (Briggs, 2005).[5] Second, and in relation to the EBM approach as a whole rather than solely meta-analysis, the role of judgement on the part of reviewers throughout the conduct of the review is contended to be of significance. Hammersley (2001) describes how procedural adherence is likely to lead to distortions in analysis, as the critical capacities of reviewers (rooted in their wider knowledge about methods and the field in question) are of greater import. These critical capacities are not transparent in any straightforward sense, but are considered essential in a critical synthesis of research. Such a synthesis should, ideally, reflect a 'skilled and knowledgeable assessment of what is likely to be true' (Hammersley, 2005, p 92); it is not 'biased' for its 'lack' of explicit procedures.

The step-by-step, sequential progression of the EBM approach is also contested on the grounds that insights may be gained in the course of conducting a review that should substantively inform the search strategy, the categories used in the analysis of data, and possibly even result in the revision of the original review question. The EBM approach does not necessarily rule out this last development (for example, see Higgins and Green, 2005, section 4.6), but doing so is likely to be seen as indicating bias, meaning that it is an option that reviewers are reluctant to take. In contrast, review methods proposed by Dixon-Woods et al (2006) and Pawson (2006) stress that although there is an overarching framework for a review that includes setting the review question, searching the literature and synthesising and analysing the data, the process is *necessarily* an iterative one. Moreover, both Dixon-Woods et al (2006) and Pawson (2006) emphasise the importance of the reviewers' critical engagement with the research being reviewed, and of a going back and forth between the primary research and the (partial) syntheses thereof rather than a single round of data extraction.

Data extraction versus data interrogation and interpretation

In the EBM approach, data are extracted from primary research in accordance with the procedure defined at the outset of the review. If the settings, populations and interventions across the various pieces of primary research are sufficiently homogeneous, then a meta-analysis will be conducted. If significant heterogeneity is present, a narrative synthesis of the data will be conducted. Whichever route is taken, the EBM approach draws a clear line between the objective presentation of an analysis (namely, the synthesis of research findings) and the interpretation of that

analysis (this being the task of, for example, policy makers) (Chalmers, 2003; Gray, 2005; Higgins and Green, 2005).

Proposed SR methods that are grounded in an interpretive tradition have a substantively different approach to the utilisation of data. Dixon-Woods et al's (2006) 'critical interpretive synthesis' method shuns formal data extraction forms in view of their resource-intensive nature and the way in which, it is contended, they inhibit flexible data extraction. Instead, a non-formalised record is utilised to inform a critical engagement with both the findings from the research and the theoretical frameworks that informed the conduct of the research. This may necessitate going back to papers to reinterpret them in the light of other work identified in the course of the review. Pawson similarly argues that the process of research synthesis and analysis is inherently interpretive and cannot be split into neat, demarcated phases where analysis and interpretation are separate; for Pawson, 'explanation-building' is the key activity in a SR, and the process of juxtaposing, reconciling, adjudicating, consolidating, and situating of 'middle-range theories' developed from the research necessarily requires an ongoing engagement with the findings and the flexibility to re-analyse other research in the light of later research (see Pawson, 2006).

Aggregative data cumulation versus the interpretive development of new knowledge

The comprehensive treatment of all data from research that meets a review's inclusion criteria is a cornerstone of the EBM approach. This comprehensiveness works in two ways. First, it is comprehensive in the sense that the review should be sensitive to the re-use of the same data (whether written up in a different form by the same authors or cited in the work of others), because to compute effect sizes without factoring this in could effectively double-count (or more) the same studies (Egger et al, 2001b). Second, it is comprehensive in the sense that the review's search strategy should not simply include flagship publications, but also less widely circulated (and non-English language) journals. The rationale here is that, because prestigious journals like to publish positive research, results showing equivocal or even harmful outcomes tend to get buried in the literature, or remain unpublished, although the evidence they contain may be of substantive import (Chalmers, 2003). The comprehensive treatment of relevant data, ideally (but not necessarily) in the form of a meta-analysis that combines the results of numerous studies (weighted according to validity if required) in order to compute a composite effect size, is thus intended to overcome the publication bias that is likely to occur if a purposive selection of data for analysis is conducted.

Proposals for SRs to take a more interpretive approach are based on a quite different view of the way in which knowledge can cumulate in a useful form for policy and practice. An interpretive approach is more concerned with developing understanding of phenomena rather than calculating an intervention's effect size. As such, it claims to be more able to deal with contextual variation as it does not seek to remove the effects of context, but rather works *with* this variability (Dixon-Woods et al, 2006; Pawson, 2006). These developments of the interpretive approach share a belief that much of the knowledge necessary for policy making and practice in complex social

contexts is not aggregative in the sense that its direction of effect can be summed. Instead, this knowledge is viewed as usable in a synthesised form only through a mental and social process that involves comparing, contrasting and debating the concepts within the research and how they may or may not transfer into other areas of policy or practice.

Conclusion

The emphasis on internal validity in the EBM approach to SRs has arguably focused debate on what is, in reality, just one component of EBP. Continuing to focus on this one area will inhibit substantive debate and methodological development. Clearly, this article does not provide any answers as to whether certain SR approaches are superior to others with regard to social policy. However, I have argued that there are reasonable grounds for believing that a strong consensus about what constitutes a scientific approach to EBP can be developed if the will to engage constructively with others advocating alternative approaches is fostered.

To this end, Donald Campbell is arguably a key figure, although perhaps not in the way that he is conventionally presumed to be. Although there are many aspects of Campbell's work that could contribute substantively to the development of SR methods, two are of particular importance. First, his disposition provides a strong example of how the research community might better approach developments and debate in the field of SRs. Throughout his working life, Campbell maintained an open mind towards new and sometimes radical approaches to research, and he relished engaging in constructive debate over the strengths and weaknesses of these approaches. Second, Campbell and his colleagues' work on how internal, external and construct validity may be attained, and when prioritising one form of validity over another may be justified, arguably provides the fundamental 'building blocks' for research (and subsequently SRs) across the field of social policy. It might also be noted that these concepts are free of the usual 'paradigmatic baggage' that can enfeeble more modern debate on SR methods.

A greater engagement with the breadth of Donald Campbell's work by those working in the field of SRs could better foster debate about different approaches to SRs for answering social policy questions. Given the complexity of his thought and the dense nature of his writing, this is not an easy task. However, if the outcome is the capability to produce richer evidence syntheses that better shape the ability of practitioners and policy makers to make effective interventions in people's lives, then the effort will be justified.

Notes

[1] For example, see the Oxford Centre for Evidence-Based Medicine's 'levels of evidence' (www.cebm.net/levels_of_evidence.asp). *Ceteris paribus*, the trustworthiness of evidence follows a clear gradation from expert opinion (weak) to systematic reviews of randomised controlled trials (strong).

[2] To date, a nomenclature for different approaches to conducting SRs has not been developed, and I have struggled here to find relevant terms. 'The dominant approach' may be suitable to a degree (any new process is likely to be initially dominated by those who spearheaded it), but the term is unhelpful for negotiating a productive path through the relevant issues. 'The traditional approach' may be more neutral, but seems ill-fitting to a process that has really only been institutionalised in the past decade. For the purposes of this article, I have decided on the inelegant but accurate 'EBM approach' to indicate the way in which EBM pioneered the use of SRs in clinical medicine and the way in which this was initially adopted as a blueprint for conducting SRs in social policy.

[3] Regarding this critique, see Byrne (2002, p 94); simplified, the law of large numbers states that, if a sample is large enough and adequately stratified, it is valid to consider the sampling distribution to be normal. *If* it is possible to draw such a sample, then Worrall's (2002) critique does not hold. Also, in a SR that utilised meta-analysis, the play of chance with respect to *unknown* confounding factors would have a decreasing effect on outcomes the greater the number of studies analysed.

[4] Research inevitably involves trade-offs; for example, random assignment can improve internal validity, but at the potential cost of failing to retain research participants who would increase the study's external validity (Shadish et al, 2002). The point being made is that the perfect piece of research is a chimera: what makes for good research is a pragmatic and judicious approach that weights internal, external and construct validity according to the particular demands for knowledge being made. See Shadish et al (2002, pp 96-102) for a full discussion of this point.

[5] A reviewer of this article correctly pointed out that there are likely to be large quantities of data whether the review is 'systematic' or not, and that formal procedures may have greater rigour where the volume of data is high. However, Briggs' (2005) critique is that the user of a meta-analysis has to largely take the results 'on trust' because of the manner in which (possibly significant) differences in study designs and contexts are subsumed in the wider analysis.

Acknowledgements

Thanks are due to Dr Ross Coomber and Professor Malcolm Williams (both at the School of Law and Social Science, University of Plymouth, UK) for providing the foundations for, and feedback on, the research that informed this article. Thanks are also due to the two anonymous referees who provided the constructive feedback on which the revision of this article was based. This work was completed with the support of an Economic & Social Research Council Collaborative Awards for Science and Engineering (CASE) postgraduate studentship (award number PTA 033 2004 00074).

References

Asthana, S. and Halliday, J. (2006) 'Developing an evidence base for policies and interventions to address health inequalities: the analysis of "public health regimes"', *Milbank Quarterly*, vol 84, no 3, pp 577-603.

Bambra, C.L. (2005) 'Reviewing the evidence: reflections from experience', *Evidence & Policy*, vol 1, no 2, pp 243-55.

Boaz, A., Ashby, D., Denyer, D., Egan, M., Harden, A., Jones, D.R., Pawson, R. and Tranfield, D. (2006) 'A multitude of syntheses: a comparison of five approaches from diverse policy fields', *Evidence & Policy*, vol 2, no 4, pp 479-502.

Bonell, C., Oakley, A., Hargreaves, J., Strange, V. and Rees, R. (2006) 'Assessment of generalisability in trials of health interventions: suggested framework and systematic review', *British Medical Journal*, vol 333, no 7563, pp 346-9.

Braye, S. and Preston-Shoot, M. (2005) 'Emerging from out of the shadows? Service users and carer involvement in systematic reviews', *Evidence & Policy*, vol 1, no 2, pp 173-93.

Briggs, D.C. (2005) 'Meta-analysis: a case study', *Evaluation Review*, vol 29, no 2, pp 87-127.

Byrne, D. (2002) *Interpreting quantitative data*, London: Sage Publications.

Cabinet Office (1999) *Professional policy making for the twenty first century*, London: Cabinet Office, www.policyhub.gov.uk/docs/profpolicymaking.pdf

Campbell, D. (1978) 'Qualitative knowing in action research', in P. Brenner, P. Marsh and M. Brenner (eds) *The social contexts of method*, London: Croom Helm.

Campbell, D. (1986) 'Science's social system of validity-enhancing collective belief change and the problems of the social sciences', in D.W. Fiske and R.A. Shweder (eds) *Metatheory in social science*, Chicago, IL: University of Chicago Press.

Campbell, D. and Russo, M.J. (1999) *Social experimentation*, London: Sage Publications.

Campbell, D. and Stanley, J. (1963) *Experimental and quasi-experimental evaluations in social research*, Chicago, IL: Rand McNally.

Chalmers, A.F. (1999) *What is this thing called science?* (3rd edition), Buckingham: Open University Press.

Chalmers, I. (2003) 'Trying to do more good than harm in policy and practice: the role of rigorous, transparent, up-to-date evaluations', *Annals of the American Academy of Political and Social Science*, vol 589, pp 22-40.

Chalmers, I. (2005) 'If evidence-informed policy works in practice, does it matter if it doesn't work in theory?', *Evidence & Policy*, vol 1, no 2, pp 227-42.

Chalmers, I. and Altman, D. (eds) (1995) *Systematic reviews*, London: BMJ Publishing Group.

Chalmers, I., Hedges, L.V. and Cooper, H. (2002) 'A brief history of research synthesis', *Evaluation and the Health Professions*, vol 25, no 1, pp 12-37.

Clarke, M. and Chalmers, I. (1998) 'Discussion sections in reports of controlled trials published in general medical journals: islands in search of continents?', *Journal of the American Medical Association*, vol 280, no 3, pp 280-2.

Cook, T.D. and Campbell, D. (1979) *Quasi-experimentation*, Chicago, IL: Rand McNally.

Davies, H. and Nutley, S. (2000) 'Healthcare: evidence to the fore', in H.T.O. Davies, S.M. Nutley and P.C. Smith (eds) *What works? Evidence-based policy and practice in public services*, Bristol: The Policy Press.

DH (Department of Health) (1999) *Saving lives: Our healthier nation*, Cm 4386, London: The Stationery Office.

Dixon-Woods, M., Cavers, D., Agarwal, S., Annandale, E., Arthur, A., Harvey, J., Hsu, R., Katbamna, S., Olsen, R., Smith, L., Riley, R. and Sutton, A.J. (2006) 'Conducting a critical interpretive synthesis of the literature on access to healthcare by vulnerable groups', *BMC Medical Research Methodology*, vol 6, no 35, www.biomedcentral.com/1471-2288/6/35

Egger, M., Davey Smith, G. and O'Rourke, K. (2001a) 'Rationale, potentials, and promise of systematic reviews', in M. Egger, G. Davey Smith and D.G. Altman (eds) *Systematic reviews in health care: Meta-analysis in context*, London: BMJ Books.

Egger, M., Dickersin, K. and Davey Smith, G. (2001b) 'Problems and limitations in conducting systematic reviews', in M. Egger, G. Davey Smith and D.G. Altman (eds) *Systematic reviews in health care: Meta-analysis in context*, London: BMJ Books.

Evidence-Based Medicine Working Group (1992) 'Evidence-based medicine: a new approach to teaching the practice of medicine', *Journal of the American Medical Association*, vol 268, no 17, pp 2420-5.

Freese, L. (1980) 'The problem of cumulative knowledge', in L. Freese (ed) *Theoretical methods in sociology*, Pittsburgh, PA: University of Pittsburgh Press.

Gallagher, D.J. (2004) 'Educational research, philosophical orthodoxy and unfulfilled promises: the quandary of traditional research in US special education', in G. Thomas and R. Pring (eds) *Evidence-based practice in education*, Maidenhead: Open University Press.

Glasziou, P., Vandenbroucke, J. and Chalmers, I. (2004) 'Assessing the quality of research', *British Medical Journal*, vol 328, no 7430, pp 39-41.

Gray, J.A.M. (2005) 'Evidence-based and value-based healthcare', *Evidence-Based Healthcare and Public Health*, vol 9, no 5, pp 317-8.

Hammersley, M. (1985) 'From ethnography to theory: a programme and a paradigm in the sociology of education', *Sociology*, vol 19, no 2, pp 244-59.

Hammersley, M. (2001) 'On "systematic" reviews of research literatures: a "narrative" response to Evans and Benefield', *British Educational Research Journal*, vol 27, no 5, pp 543-54.

Hammersley, M. (2004) 'Some questions about evidence-based practice in education', in G. Thomas and R. Pring (eds) *Evidence-based practice in education*, Maidenhead: Open University Press.

Hammersley, M. (2005) 'Is the evidence-based practice movement doing more good than harm? Reflections on Iain Chalmers' case for research-based policy making and practice', *Evidence & Policy*, vol 1, no 1, pp 85-100.

Harden, A. (2001) 'The fine detail: conducting a systematic review', in S. Oliver and G. Peersman (eds) *Using research for effective health promotion*, Buckingham: Open University Press.

Haynes, R.B. (2005) 'Of studies, summaries, synopses, and systems: the "4s" evolution of services for finding current best evidence', *Evidence-Based Nursing*, vol 8, no 1, pp 4-6.

Higgins, J.P.T. and Green, S. (2005) *Cochrane handbook for systematic reviews of interventions: version 4.2.5* (updated May 2005), Chichester: John Wiley and Sons.

Holmes, D., Murray, S.J., Perron, A. and Rail, G. (2006) 'Deconstructing the evidence-based discourse in health sciences: truth, power and fascism', *International Journal of Evidence Based Healthcare*, vol 4, no 3, pp 180-6.

Home Office (1999) *Reducing crime and tackling its causes: A briefing note on the Crime Reduction Programme*, London: Home Office.

Killoran, A. and Kelly, M. (2004) 'Towards an evidence-based approach to tackling health inequalities: the English experience', *Health Education Journal*, vol 63, no 1, pp 7-14.

Knorr-Cetina, K.D. (1981) *The manufacture of knowledge: An essay on the constructivist and contextual nature of science*, Oxford: Pergamon Press.

Kuhlmann, E. (2004) 'Knowledge and power in the era of consumerism – negotiations on professionalism in health care', Paper for presentation at the International Sociological Association interim conference 'RC52, Knowledge, Work and Organization', 22-24 September, Versailles, www.zes.uni-bremen.de/~kuhlmann/ISA04_kuhlmann.pdf

Latour, B. and Woolgar, S. (1986) *Laboratory life: The construction of scientific facts* (2nd edition), Princeton, NJ: Princeton University Press.

Littlejohns, P. and Chalkidou, K. (2006) 'Evidence for health policy', *Journal of Research in Nursing*, vol 11, no 2, pp 110-17.

Oakley, A. (2000) *Experiments in knowing: Gender and method in social sciences*, Cambridge: Polity Press.

Oakley, A. (2006) 'Resistances to "new" technologies of evaluation: education research in the UK as a case study', *Evidence & Policy*, vol 2, no 1, pp 63-87.

Oakley, A., Gough, D., Oliver, S. and Thomas, J. (2005) 'The politics of evidence and methodology: lessons from the EPPI-centre', *Evidence & Policy*, vol 1, no 1, pp 5-31.

Oliver, S., Harden, A., Rees, R., Shepherd, J., Brunton, G., Garcia, J. and Oakley, A. (2005) 'An emerging framework for including different types of evidence in systematic reviews for public policy', *Evaluation*, vol 11, no 4, pp 428-46.

Pawson, R. (2002) 'Evidence-based policy: the promise of "realist synthesis"', *Evaluation*, vol 8, no 3, pp 340-58.

Pawson, R. (2006) *Evidence-based policy: A realist perspective*, London: Sage Publications.

Pawson, R. and Tilley, N. (1997) *Realistic evaluation*, London: Sage Publications.

Petticrew, M. and Roberts, H. (2006) *Systematic reviews in the social sciences: A practical guide*, Oxford: Blackwell.

Sackett, D.L., Richardson, W.S., Rosenberg, W. and Haynes, R.B. (1997) *Evidence-based medicine: How to practice and teach EBM*, London: Churchill Livingstone.

Sackett, D.L., Rosenberg, W.M.C., Gray, J.A.M., Haynes, R.B. and Richardson, W.S. (1996) 'Evidence-based medicine: what it is and what it isn't', *British Medical Journal*, vol 312, no 7023, pp 71-2.

Shadish, W.R., Cook, T.D. and Campbell, D. (2002) *Experimental and quasi-experimental designs for generalized causal inference*, Boston: Houghton Mifflin Company.

Sharland, E. and Taylor, I. (2006) 'Social care research: a suitable case for systematic review?', *Evidence & Policy*, vol 2, no 4, pp 503-23.

Slavin, R.E. (1995) 'Best-evidence synthesis: an intelligent alternative to meta-analysis', *Journal of Clinical Epidemiology*, vol 48, no 1, pp 9-18.

Victora, C.G., Habicht, J.-P. and Bryce, J. (2004) 'Evidence-based public health: moving beyond randomized trials', *American Journal of Public Health*, vol 94, no 3, pp 400-8.

Wallace, A., Croucher, K., Bevan, M., Jackson, K., O'Malley, L. and Quilgars, D. (2006) 'Evidence for policy making: some reflections on the application of systematic reviews to housing research', *Housing Studies*, vol 21, no 2, pp 297-314.

Weightman, A., Ellis, S., Cullum, A., Sander, L. and Turley, R. (2005) *Grading evidence and recommendations for public health interventions: Developing and piloting a framework*, London: Health Development Agency, www.nice.org.uk/page.aspx?o=503421

Worrall, J. (2002) '*What* evidence in evidence-based medicine?', *Philosophy of Science*, vol 69, no 3 (supplement), pp S316-S330.

Mark Pearson, School of Law and Social Science, University of Plymouth, UK, mark.pearson@plymouth.ac.uk