2018-06-05

Supplementary search methods were more effective and offered better value than bibliographic database searching: a case study from public health and environmental enhancement.

Cooper, C

10.1002/jrsm.1286
Research Synthesis Methods
Wiley

Full Title: supplementary search methods were more effective and offered better value than bibliographic database searching: a case study from public health and environmental enhancement.

Short Title: supplementary versus databases: a case study.

Names and affiliations of authors:

(Corresponding author): Chris Cooper. PenTAG, University of Exeter Medical School. St Luke's Campus, Exeter. EX1 2LU e: Christopher.Cooper@exeter.ac.uk t: 01392 726095 (we don't have a fax machine so number not listed).

Rebecca Lovell, European Centre for Environment and Human Health, University of Exeter Medical School, Truro Campus, Truro, TR1 3HD: r.lovell@exeter.ac.uk , 01872 258 173

Kerryn Husk, NIHR CLAHRC South West Peninsula (PenCLAHRC), Plymouth University Peninsula Schools of Medicine and Dentistry, N10, ITTC Building, Plymouth Science Park, Plymouth, UK. PL6 8BX

Andrew Booth. HEDS, School of Health and Related Research (ScHARR), University of Sheffield.

Ruth Garside
European Centre for Environment and Human Health, University of Exeter Medical School, Truro.

This research and the original systematic review were funded by the National Institute for Health Research School for Public Health Research (NIHR SPHR) Programme. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health. The School for Public Health Research (SPHR) is funded by the National Institute for Health Research (NIHR). SPHR is a partnership between the Universities of Sheffield, Bristol, Cambridge, UCL; The London School for Hygiene and Tropical Medicine; The Peninsula College of Medicine and Dentistry; the LiLaC collaboration between the Universities of Liverpool and Lancaster and Fuse; The Centre for Translational Research in Public Health, a collaboration between Newcastle, Durham, Northumbria, Sunderland and Teesside Universities. The European Centre for Environment and Human Health (part of the University of Exeter Medical School) is part financed by the European Regional Development Fund Programme 2007 to 2013 and European Social Fund Convergence Programme for Cornwall and the Isles of Scilly.

CC is a part-time PhD student and he is supported by the NIHR Health Technology Assessment Grant. This study forms a part of his PhD thesis.

KH and RG are supported by the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care South West Peninsula. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

Background: We undertook a systematic review to evaluate the health benefits of environmental enhancement and conservation activities. We were concerned that a conventional process of study identification, focusing on exhaustive searches of bibliographic databases as the primary search method would be ineffective, offering limited value.

The focus of this study is comparing study identification methods. We compare: (i) an approach led by searches of bibliographic databases to (ii) an approach led by supplementary search methods. We retrospectively assessed the effectiveness and value of both approaches.

Methods: 'Effectiveness' was determined by comparing: 1) the *total number of studies* identified and screened and, 2) the number of includable studies *uniquely identified* by each approach.

'Value' was determined by comparing included study quality and by using qualitative sensitivity analysis to explore the contribution of studies to the synthesis.

Results: The bibliographic databases approach identified 21,409 studies to screen and two included qualitative studies were uniquely identified. Study quality was moderate and contribution to the synthesis was minimal.

The supplementary search approach identified 453 studies to screen and nine included studies were uniquely identified. Four quantitative studies were poor quality but made a substantive contribution to the synthesis; Five studies were qualitative: three studies were good quality, one was moderate quality, and one study was excluded from the synthesis due to poor quality. All four included qualitative studies made significant contributions to the synthesis.

Conclusions: This case study found value in aligning primary methods of study identification to maximise location of relevant evidence.

Keywords: information science; literature searching; sensitivity analysis; Cochrane systematic reviews; Public health.

Acknowledgments: Jo Varley Campbell and Danica Cooper for proof-reading the final draft and their help in collation of tables.

Background

With the increased interest in evidence-informed environmental policy (Dicks et al., 2014), researchers have explored the suitability of applying the explicit methods of systematic review to the field of conservation research (Pullin and Knight, 2001, Fazey et al., 2004, Stewart et al., 2005, Haddaway and Bayliss, 2015, Bilotta et al., 2014b, Bilotta et al., 2014a). Whilst collectively researchers agree that a systematic process to identify and review studies is of benefit, they helpfully highlight several issues. A primary concern is the appropriateness and application of a process and methodology which was originally developed to systematically review studies reporting randomised controlled trials indexed within bibliographic databases, to the systematic review of the myriad of study designs used to evaluate conservation, and other complex interventions, the results of which are widely dispersed throughout academic databases and 'grey literature' (Pullin and Knight, 2001, Fazey et al., 2004, Stewart et al., 2005).

In 2012, we began a mixed-methods systematic review to evaluate the health and wellbeing impacts for different groups of people undertaking environmental enhancement and conservation activities (NIHR, 2012). We encountered issues highlighted by Pullin and Knight, Fazey et al, and Stewart et al (Pullin and Knight, 2001, Fazey et al., 2004, Stewart et al., 2005) as we began scoping our review, namely: a relative absence of studies using controlled or otherwise 'higher order' study designs (Stewart et al., 2005, Fazey et al., 2004, Haddaway and Bayliss, 2015); a difficulty in accessing primary studies to review, due to: delays in publication, limited publication, or simply no attempt to formally publish completed research (Kareiva et al., 2002, Haddaway and Bayliss, 2015); and a recognition that a variety of sources would need to be searched to identify studies (Fazey et al., 2004, Kareiva et al., 2002). Our project reference group (PRG[1]) validated these concerns, while anticipating that many of the studies that might address our research question, would likely be found in the grey literature.

We were concerned that a conventional approach to study identification, described in the leading handbooks for the process of systematic review (LEFEBVRE, 2011, Centre for Reviews and Dissemination and (CRD), 2009) that focuses on sensitive searches of bibliographic databases as the primary method of study identification, could yield an overwhelming number of studies to screen, with low numbers of includable studies identified, and potentially diverting time away from identification of grey literature. Facing similarly challenging searches, other researchers have explored the successful adaptation of conventional search

[1] practitioners, experts in the field and academics brought together to oversee the development of the review

methods to the identification of studies within disparate bodies of grey literature (Adams et al., 2016, Godin et al., 2015, Mahood et al., 2014). Accordingly, we developed a tailored study identification protocol. The tailored study identification protocol was designed *a priori* to ensure the systematic identification of studies and minimise the introduction of bias in study selection, whilst also seeking to allocate time to supplementary study identification methods that were anticipated to offer a more productive yield of studies for inclusion than searches of bibliographic databases.

During the process of protocol development, we registered our systematic review with Cochrane's Public Health Group (Husk et al., 2013). Cochrane provides specific methodological guidance for the systematic review of intervention effectiveness. Typically, in Cochrane reviews of interventions, studies reporting randomised controlled trials are sought (LEFEBVRE, 2011) but, in public health reviews and/or reviews of conservation interventions such as this one, a range of study designs may be included (Armstrong R, 2011). The process of study identification for Cochrane Reviews is set out in detail in chapter six of The Cochrane Handbook, 'searching for studies,' and summarised for reviews in public health topics in chapter 21, 'reviews in public health and health promotion' (LEFEBVRE, 2011, Armstrong R, 2011). The aim of study identification within the Cochrane model is the comprehensive identification of published and unpublished studies; this is a sequential process of study identification, led by comprehensive searches of bibliographic databases and followed by searches of non-bibliographic databases sources (e.g. handsearching, searches of conferences).

As Cochrane authors, we were committed to following this Cochrane process of study identification but, given the need to interpret this process within conservation science and public health, and our awareness of the need for more time and effort to identify grey literature than is typical for a Cochrane review, we decided to employ a hybrid approach. This augmented the Cochrane method for study identification (with bibliographic database searches as its primary method of study identification) with a tailored study identification protocol (with supplementary searches as its primary method of study identification and a focus on extensive grey literature searches). This adaptation provided us with the opportunity to compare the effectiveness of the two study identification protocols.

Study aims

To assess the effectiveness and value of a search approach led by supplementary search methods (the tailored study identification protocol) compared to a search approach led by bibliographic databases (The Cochrane study identification protocol).

In this study, we determined 'effectiveness' by comparing (i) the *total number of studies* identified and screened and (ii) by comparing the number of included studies *uniquely identified* by each study identification protocol. We determined 'value' by comparing the study quality across included studies retrieved for each study identification protocol and by analysing the contribution of studies to the synthesis.

Developing the Cochrane study identification protocol and tailored study identification protocol

This section describes how we developed the Cochrane study identification protocol and the tailored study identification protocol and the methods used to measure the effectiveness of study identification and the evaluation of study quality and contribution to the synthesis of each approach.

The Cochrane study identification protocol

The Cochrane study identification protocol was developed and peer-reviewed as a required component of our overall systematic review protocol by The Cochrane Public Health Group (Husk et al., 2013).

The primary method of study identification in the Cochrane study identification protocol involved searches of 22 bibliographic databases (see figure four). The multi-disciplinary nature of conservation/public health topics means that studies can be identified from diverse databases, not necessarily limited to health topics, so it is common practice to search a greater number of bibliographic databases than for clinical topics (Beahler et al., 2000, Grayson L, 2003, Bayliss and Beyer, 2015, Bayliss et al., 2014). These 22 databases included: MEDLINE (OVID), Embase (OVID) and The Cochrane Library (Wiley interface) as well as Social Policy and Practice (OVID), IBSS (Pro Quest) and ASSIA (Pro Quest), CAB Abstracts and Greenfile. The full list of bibliographic databases searched, and our MEDLINE search strategy, is included in the published Cochrane review (Husk et al., 2016). The Trial Search Co-Ordinator of The Cochrane Public Health Group checked and approved our searches.

The tailored study identification protocol

The tailored study identification protocol included the same methods of study identification as set out in The Cochrane Handbook (and used in the Cochrane protocol) but with a revised focus for study identification methods. We changed the primary focus of study identification from bibliographic database searching to contacting organisations and searching web-sites thereby affecting the weighting of the methods in the process of study identification as it relates to searching time. Studies evaluating the use of supplementary search methods were useful in informing this discussion (Papaioannou et al., 2010).

The study identification protocols are outlined in figure one.

The design of the tailored study identification protocol

We sought to sensitise the team to the disparate evidence for this review before designing the tailored study identification protocol. We aimed to understand what types of studies (by design, publication type and publication status) may exist and where (and how) they could be identified. We sought to achieve this in two ways:

1. scoping searches were undertaken by the review team. Scoping searches took the following structure: ((search terms for possible interventions) and (search terms for review-relevant outcomes)). The aim was to identify candidate studies in bibliographic databases (published) and through web-searching (grey literature). The purpose of these searches was early identification of studies and organisations as well as to explore how and where potentially includable studies were being identified; and
2. a project reference group (**PRG**) was formed, made up of a wide range of key organisations, such as: the Conservation Volunteers, Mind, Local Authorities and Groundwork. We met with the PRG at a preliminary stage in our review to hear from topic experts about the types of interventions and participants we were aiming to find/identify. This helped generate search terms and it developed our understanding of the evidence base for the review, in particular the nature of the grey literature.

Whilst the process described above was iterative and informal, it identified two key factors that ultimately informed the order of study identification methods in the tailored study identification protocol. First, the PRG advised that the types of studies that would meet our inclusion criteria were likely to be identified in the grey literature and, secondly, our scoping searches of bibliographic databases suggested that a sensitive search strategy for this review

would yield approximately 20,000 studies to screen. Piloting our inclusion/exclusion criteria on these 20,000 studies suggested low specificity and precision suggesting the need to prioritise grey literature searches as a way to further refine the bibliographic database search strategy.

The tailored study identification protocol was designed therefore to concentrate searching time on grey literature searches as the primary method of study identification, specifically contacting organisations and experts in the field to identify studies, supplemented with web searching. In contrast to the Cochrane study identification protocol, we planned that bibliographic database searching would be a supplementary search method to identify published studies and reviews.

Methods

This is a retrospective comparison of the effectiveness and value of the two study identification protocols.

Effectiveness

Effectiveness is a term used in literature searching to describe the impact of study identification when two (or more) search approaches are compared. Whilst methods exist to calculate search effectiveness (e.g. sensitivity, specificity and precision), there is no agreed understanding as to what actually constitutes effectiveness in study identification. In this study 'effectiveness' will be determined by: 1) comparing the *total number of studies* identified and screened by each of the two study identification protocols and 2) comparing the number of included studies *uniquely identified* by each of the two study identification protocols. We are able to make this comparison since the same inclusion and exclusion criteria were used to screen studies returned by each study identification protocol.

Value and contribution

Determining effectiveness in purely quantitative terms as the number of studies identified and included in the review (as above) makes no acknowledgement of the *value* of the studies identified uniquely by each study identification protocol, nor how studies may substantively *contribute* to the synthesis or alter the conclusions of the review. In this study, we seek to link the idea of effectiveness (defined above) to the concept of study value (defined below), so that we can determine not only the effect of each study identification protocol but also the value. Value will be determined by comparing a measure of study 'quality' and by assessing the unique contribution from each study identified to the synthesis and the confidence in the findings.

Study quality

The assessment of study 'quality', using standardised and validated tools, is a key component in a systematic review (Garside, 2014). Quality assessment of studies included in a review examines the risk of bias in studies using quantitative study designs, and subjective interpretation in qualitative studies, and the impact on results (Sterne JAC et al., 2011), guiding the interpretation of findings (Armijo-Olivo et al., 2012). In this way, study quality is integral to interpreting the value of studies identified.

Study quality was assessed using the Effective Public Health Practice Project (EPHPP) tool for studies using quantitative study designs (Effective Public Health Practice Project, 2009). Study quality was rated over six categories from being very strong (scoring the minimum of 6) up to very weak (scoring the maximum of 18). Scoring for these six categories where, 1 = strong, 2 = moderate and 3 = weak.

Cochrane's risk of bias tool was not used in the absence of any includable RCTs (Husk et al., 2013). The Wallace criteria were used to appraise qualitative studies (Wallace et al., 2004).

Contribution to the synthesis (qualitative studies only)

We are not aware of any formal or standardised approach to identifying the 'contribution' of any individual study to the findings in a qualitative synthesis, although researchers describe the use of 'sensitivity analysis' (Thomas and Harden, 2008). We developed an alternative approach and we test this idea here for the first time in an attempt to link methods for study identification to study value.

Contribution to the synthesis was evaluated by re-examining the qualitative synthesis (e.g. the documentation of the results of each of the individual stages of the qualitative synthesis) to understand which papers substantively contributed data, concepts and understanding to identification and development of the overarching themes and sub-themes. The synthesis of qualitative studies as reported in our Cochrane review was used (Husk et al., 2016). Once each paper's contribution to the overarching and sub-themes was identified in the synthesis, we determined which studies were: 1) fundamental and necessary to the specific overarching and/or sub-theme (we term these 'key studies'), and 2) which papers merely added confirmatory validity or data richness (we term these 'additional studies'). This contributed an understanding of the relative contribution of each paper to the overall synthesis. The Confidence in the Evidence from Reviews of Qualitative Research (CERQual) approach was then used to appraise the confidence in review findings with and without the studies that were missed by each study identification protocol (Lewin et al., 2015). The CERQual tool helps assess how much confidence to place in the findings from a qualitative evidence synthesis (Lewin et al., 2015). In this study, we make the link between confidence and attempt to interpret this as value.

Results

Effectiveness

The number of studies identified and screened by each study identification protocol

The Cochrane study identification protocol resulted in the identification of 21,409 studies to screen at the title/abstract stage, compared with 453 studies identified via the tailored study identification protocol searches. At full text, 166 studies were screened from the Cochrane study identification protocol and 211 were screened from the tailored study identification protocol

The number of studies uniquely identified by each study identification protocol

Twenty-one studies met our review inclusion criteria and were included in the review (figure two). By study identification protocol these were:

Studies identified by the Cochrane study identification protocol only: two

Two included studies were uniquely identified by the Cochrane study identification protocol through bibliographic database searching (Burls, 2007, Gooch, 2005) (figure 2). Burls et al (Burls, 2007) was identified twice: once in Social Policy and Practice (OVID) and again in British Nursing Index (Pro Quest). Gooch et al (Gooch, 2005) was identified once, in the International Bibliography of the Social Sciences (IBSS, Pro Quest).

Studies identified by the tailored study identification protocol only: nine

Nine included studies were uniquely identified by the tailored study identification protocol (figure 2) (Brooker and Brooker, 2008a, Brooker and Brooker, 2008b, BTCV, 2010, Christie, 2004, Eastaugh et al., 2010, Halpenny and Caissie, 2003, Small Woods, 2011, Wilson,

2009, Yerrell, 2008). These studies were uniquely identified by the tailored study identification protocol and were not indexed in any of the bibliographic databases. These studies could only have been identified by author contact or web-searching.

*Study identified by citation chasing (Cochrane study identification protocol **and** tailored study identification protocols): one*

One included study was identified uniquely by citation chasing, a method of study identification shared by both search protocols (figure 2). Townsend et al (Townsend and Marsh, 2004) was identified through backwards citation chasing Moore et al which was identified by both search protocols (Moore et al., 2006).

Studies identified by both study identification protocols: nine

Nine included studies were identified by both the tailored protocol and the Cochrane protocol (figure 2) (Barton, 2009, Birch, 2005, Carter, 2008, O'Brien et al., 2010, O'Brien et al., 2008, Pillemer, 2010, Reynolds, 1999, Townsend, 2006, Townsend and Moore, 2005). These studies were identified by bibliographic searching in the Cochrane study identification protocol and, separately, through organisation contact and web-searching in the tailored study identification protocol.

Effectiveness summary

The tailored study identification protocol identified all but two studies: a study by Burls and a study by Gooch, both qualitative studies (Burls, 2007, Gooch, 2005). The tailored study identification protocol uniquely identified nine studies missed by the Cochrane study identification protocol (Brooker and Brooker, 2008a, Brooker and Brooker, 2008b, BTCV, 2010, Christie, 2004, Eastaugh et al., 2010, Halpenny and Caissie, 2003, Small Woods, 2011, Wilson, 2009, Yerrell, 2008).

Value

Study quality

Quantitative studies: The EPHPP Tool

The EPHPP tool scores study quality using a global rating summarised in three domains: Strong, Moderate and Weak (Effective Public Health Practice Project, 2009). The tailored study identification protocol uniquely identified seven studies using quantitative study designs and the quality was scored weak for all (between 12-18. Table 1). Two of these seven studies were included in our review but were excluded from the actual synthesis due to poor study quality (primarily due to small study samples) (Brooker and Brooker, 2008a, Brooker and Brooker, 2008b). No studies using quantitative study designs were identified uniquely by the Cochrane study identification protocol (Table 1).

Qualitative studies: The Wallace Criteria

Where seven or more of the Wallace criteria were answered positively, studies were scored as 'good', if studies met between four and six criteria positively, a 'moderate' score was awarded.

In total, nine qualitative studies were identified (Table 1). The two studies uniquely identified by the tailored study identification protocol were scored as 'good' (Christie, 2004, Halpenny and Caissie, 2003) whereas the two studies uniquely identified by the Cochrane study identification protocol were scored as 'moderate' (Burls, 2007, Gooch, 2005). This data, and the quality appraisal of the studies identified by both the tailored study identification protocol and the Cochrane study identification protocol, is set out in Table 1.

Contribution to synthesis

The contributions of the quantitative and qualitative studies have been appraised separately. For the mixed method studies, these studies (Wilson 2009, Yerrell 2008 and O'Brien 2008) have been appraised separately for their contributions of quantitative and qualitative data.

Quantitative

No studies reporting quantitative data were uniquely identified by the Cochrane study identification protocol so the results reported here focus on the seven studies uniquely identified by the tailored study identification protocol and the five studies identified by both protocols. The heterogeneity of outcomes assessed by the study authors, the general lack of studies using controlled study designs, and the poor study quality overall, prohibited meta-analysis. The results are therefore summarised narratively and tabulated in Table 2 below.

Five outcome domains were of interest in this review:

1. physiological outcomes,
2. physical health measures,
3. mental and emotional wellbeing,
4. quality of life, and
5. physical activity measures

The tailored study identification protocol identified studies that contributed data to three of these outcomes: mental and emotional wellbeing (Wilson, 2009); quality of life (BTCV, 2010, Eastaugh et al., 2010, Small Woods, 2011, Wilson, 2009, Yerrell, 2008) and physical activity measures (Wilson, 2009).

In the first domain (mental and emotional wellbeing), the identification and inclusion of Wilson et al did not alter the overall conclusion of improvements of mental and emotional wellbeing (Wilson, 2009, Husk et al., 2013).

In the second domain (quality of life), one study reported HRQoL improvements (Yerrell, 2008). Two studies also reported improvements in HRQoL, one from the tailored study identification protocol (Small Woods, 2011) and another identified by the tailored study identification protocol and the Cochrane study identification protocol (Reynolds, 1999), but both studies had small sample sizes (Small Woods n=7 & Reynolds n=15 compared with Yerrell n=194) which limits the robustness of the findings (Husk et al., 2013). The findings of Yerrell would therefore appear valuable in this domain, in relation to their findings and relative to their sample size, although the uncontrolled before-and-after study design is considered of limited value in assessing causation (Yerrell, 2008, Husk et al., 2013).

One study was unique to the tailored study identification protocol in the final domain (physical activity measures) (Wilson, 2009). Wilson et al reported increased physical activity, measured using a validated tool,12 weeks after participating in environmental enhancement activities (Wilson, 2009). Only one other study evaluated physical activity measures (Pillemer, 2010). The study by Pilemer, identified by both the tailored and the Cochrane study identification protocols, also found improvements in physical activity scores but this was appraised retrospectively and through a scale created especially for their study (Pillemer, 2010). The findings of Wilson et al would therefore appear valuable in this domain (Wilson, 2009, Husk et al., 2013).

Quantitative summary

Whilst the quality of each study (and therefore of the overall pool of studies) was weak regardless of study identification protocol, the value of each of the studies to the synthesis is clear. To generate a reliable understanding of intervention effectiveness, it was important that all studies reporting effectiveness outcomes are identified and the Cochrane study identification protocol would have missed studies and, thus, study data.

Qualitative

The findings of the qualitative studies were used to understand the links, as perceived by participants, between participation in environmental enhancement activities and health and wellbeing outcomes (Lovell et al., 2015, Husk et al., 2016).

Nine overarching themes were identified in the qualitative synthesis:

1. Physical activity
2. Personal achievement
3. Personal/ social identity
4. Developing knowledge
5. Benefits of place
6. Social Contact
7. Spirituality
8. Psychological benefits
9. Risks/negatives

Evidence available per theme

Table 3 records the study data available per theme. Eight of the nine themes were present in one or more of the studies rated as 'good' quality (Table 1) (Lovell et al., 2015).

Contribution of studies per theme

The results of the analysis to determine the contribution of individual studies to the synthesis are recorded below. The first theme, Physical Activity, is summarised narratively and through figure three. The remaining eight themes are summarised narratively but with the corresponding figures being included in the supplementary file.

Studies are categorised as 'key studies' where they provide sufficient validity and richness to identify key concepts and develop primary and sub-themes. If a study provides either data richness, through a participant quotation to support a sub-theme, or a study confirms validity through identifying the themes and being cited in the final review, we categorise this as an 'additional study' since it provides additional but not unique contributions. If a study is identified as a 'key study' but it is also an additional study for another sub-theme, it is only counted once as a key study in the narrative since the synthesis is dependent on it.

Physical activity

Figure three summarises the contribution of studies to this theme. Overall seven studies contributed data to this theme. Analysis of the sub-themes shows that five of the seven studies were 'key studies' with sufficient validity and richness to identify key concepts and develop primary and sub-themes (Townsend, 2006, Townsend and Marsh, 2004, O'Brien et al., 2008, BTCV, 2010, Carter, 2008, Wilson, 2009). Two studies provided data that reinforced the primary theme or sub-themes identified from the key studies but did not contribute new knowledge to the synthesis (Burls, 2007, Birch, 2005).

Personal achievement (see supplementary file 2 for summary figure)

Overall, twelve studies contributed data to this theme. Analysis of the sub-themes shows that two studies were 'key studies' with sufficient validity and richness to identify all key concepts and develop primary and sub-themes (Wilson, 2009, Christie, 2004). Five studies provided data that reinforced the primary theme or sub-themes identified from the key studies but did not contribute new knowledge to the synthesis (BTCV, 2010, Burls, 2007, Gooch, 2005, Townsend, 2006, Townsend and Marsh, 2004).

Personal/ social identity

Overall, six studies contributed data to this theme. Analysis of the sub-themes shows that three of the five studies were 'key studies' with sufficient validity and richness to identify key concepts and develop primary and sub-themes (Carter, 2008, Christie, 2004, O'Brien et al., 2008). Three studies provided data that supported the primary theme or sub-themes identified from the key studies but did not contribute new knowledge to the synthesis (Gooch, 2005, Wilson, 2009, Burls, 2007).

Developing knowledge

Overall, nine studies contributed data to this theme. Analysis of the sub-themes shows that three of the nine studies were 'key studies' with sufficient validity and richness to identify key concepts and develop primary and sub-themes (O'Brien et al., 2010, O'Brien et al., 2008, BTCV, 2010). Six studies provided data that supported the primary theme or sub-themes identified from the key studies but did not contribute new knowledge to the synthesis (Burls, 2007, Gooch, 2005, Wilson, 2009, Halpenny and Caissie, 2003, Townsend, 2006, Christie, 2004, Carter, 2008).

Benefits of place

All 12 studies contributed data to this theme. Analysis of the sub-themes shows that five studies were 'key studies' with sufficient validity and richness to identify all key concepts and develop primary and sub-themes (O'Brien et al., 2008, Townsend and Marsh, 2004, Halpenny and Caissie, 2003, Christie, 2004, Wilson, 2009). Two studies provided data that supported the primary theme or sub-themes identified from the key studies but did not contribute new knowledge to the synthesis (Gooch, 2005, Burls, 2007).

Social contact

All 12 studies contributed data to this theme. Analysis of the sub-themes shows that five studies were 'key studies' provided sufficient validity and richness to identify all key concepts and develop primary and sub-themes (BTCV, 2010, O'Brien et al., 2010, O'Brien et al., 2008, Carter, 2008, Halpenny and Caissie, 2003). One study provided data that supported the primary theme or sub-themes identified from the key studies but did not contribute new knowledge to the synthesis (Gooch, 2005).

Spirituality

Overall, five studies contributed data to this theme. Analysis of the sub-themes shows that two studies were key studies with sufficient validity and richness to identify all key concepts and develop the primary theme and sub-themes (O'Brien et al., 2010, Christie, 2004). Three studies provided data that supported primary or sub-themes identified from the key studies but did not contribute new knowledge to the synthesis (O'Brien et al., 2008, Burls, 2007, BTCV, 2010).

Psychological benefits

Overall, eleven studies contributed data to this theme. Analysis of the sub-themes shows that two studies were key studies with sufficient validity and richness to identify key concepts and develop the primary theme and sub-themes (Wilson, 2009, Christie, 2004). Three studies provided data that supported primary or sub-themes identified from the key studies but did

not contribute new knowledge to the synthesis (Halpenny and Caissie, 2003, Gooch, 2005, Birch, 2005, Burls, 2007).

Risk and negative impacts

Overall, four studies contributed data to this them. Analysis of the sub-themes shows that one of the five studies provided sufficient validity and richness to identify key concepts and develop primary and sub-themes (Christie, 2004). Two studies provided data that supported the primary theme or sub-themes identified from the key studies but did not contribute new knowledge to the synthesis (Burls, 2007, Gooch, 2005).

Qualitative summary

Within the nine overarching themes, 37 sub-themes were identified from nine studies (Townsend, 2006, Townsend and Marsh, 2004, O'Brien et al., 2008, BTCV, 2010, Carter, 2008, Wilson, 2009, Halpenny and Caissie, 2003, O'Brien et al., 2010, Christie, 2004). These nine studies were fundamentally key to the synthesis since they provided sufficiently rich data to identify key concepts and develop all the overarching themes and sub-themes. If any of these studies had been missed, the findings of the review would have been different since potentially unique data from sufficiently rigorous studies would have been omitted from the synthesis. The identification and contribution of these nine studies was therefore key to the qualitative review. These nine studies were all identified by the tailored study identification protocol.

Studies supporting either overarching or sub-themes were included in the synthesis. Whilst the identification and inclusion of these studies increase the validity of the overall synthesis, two studies were only used in the synthesis to increase validity and they did not identify primary or sub-themes uniquely (Burls, 2007, Gooch, 2005, Birch, 2005). The omission of these studies from the synthesis would not alter the synthesis or change the findings of the review. These studies were uniquely identified by the Cochrane study identification protocol (Burls, 2007, Gooch, 2005).

The CERQual tool was used to appraise how much confidence could be placed in the findings listed above and its application in this study extends the work undertaken in our Cochrane Review. In this study, we first applied CERQual to all findings and included all studies in the analysis (table 4). Secondly, we applied CERQual to all findings but excluded the study by Burls and the study by Gooch, since we sought to measure the contribution of bibliographic database searching in the Cochrane study identification protocol and the potential impact of missing these studies on the synthesis of studies (table 5). Thirdly, we applied CERQual to all findings but excluded the study by Christie and the study by Halpenny and Cassie, since we sought to measure the contribution of author contact in the tailored protocol and the potential impact of missing these studies on the synthesis of studies (table 6).

The use of CERQual allows us to measure the impact of potentially missing studies from either search protocol and to explore any possible changes to the synthesis of studies. It also helps demonstrate the utility of both search approaches, helping us to interpret the value of studies and, therefore, the search protocols or search methods.

CERQual: excluding the study by Burls and the study by Gooch (table 5) We found no difference in the overall confidence of findings in any of the nine domains if the study by Burls and the study by Gooch were removed. We observed small changes in the assessment of adequacy in three cases but these changes did not alter the overall confidence using CERQual. These changes were:

- physical activity: minor methodological limitations were consistent between both analyses. This did not change the overall CERQual assessment of moderate confidence;
- personal achievement: the removal of Burls raised minor concerns in the assessment of adequacy but the overall CERQual assessment of high confidence remained unchanged;
- social contact: the use of Gooch to provide validating richness was a minor concern in the assessment of adequacy but the overall CERQual assessment of high confidence remained unchanged; and
- risks and negative impacts: minor methodological limitations were noted in the assessment of adequacy, since the removal of Gooch would potentially remove a sub-theme. This would not, however, change the overall CERQual assessment of moderate confidence in this domain. Overall, this domain was of limited importance to the synthesis.

This analysis would appear to confirm our finding that the study by Burls and the study by Gooch did not materially affect the synthesis of qualitative studies. This would suggest that in missing these particular studies the synthesis, as presented in our Cochrane review, would remain unchanged.

CERQual: excluding the study by Christie and the study by Halpenny & Cassie (table 6)
We observed a difference in the overall confidence of findings in five of the nine domains if the study by Christie and the study by Halpenny & Cassie were removed. These changes significantly altered the confidence in findings and, therefore, would appear to impact negatively on the synthesis of studies had these two studies been missed by our searches. The changes were in the following domains:

- personal achievement: the CERQual assessment was altered by the removal of these two studies, being downgraded from high confidence to moderate confidence. The loss of Christie (specifically) raised major concerns in the assessment of adequacy and minor concerns in the assessment of coherence. Furthermore, minor concerns were raised in methodological limitations, since both the removed studies were 'good quality' studies;
- personal/social identity: the CERQual assessment was altered by the removal of these two studies, being downgraded from high confidence to moderate confidence. The loss of Christie raised concerns on adequacy and coherence specifically;
- developing knowledge: there was no change in the CERQual assessment. This theme was graded as high confidence even in spite of the omission of Christie;
- benefits of place: the CERQual assessment was altered by the removal of Christie, being downgraded from high confidence to moderate confidence. The loss of Christie raised concerns on adequacy specifically;
- social contact: the CERQual assessment was altered by the removal of these two studies, being downgraded from high confidence to moderate confidence;
- spirituality: the CERQual assessment was altered by the removal of Christie, being downgraded from high confidence to low confidence. The loss of Christie raised concerns on adequacy; and
- risks and negative impacts: minor methodological limitations were noted in the assessment of adequacy. This would not, however, change the overall CERQual assessment of moderate confidence in this domain. Overall, this domain was of limited importance to the synthesis.

This additional analysis would appear to confirm our finding that the study by Burls and the study by Gooch did not materially affect the synthesis of qualitative studies, whereas the studies by Christie and Halpenny and Cassie did.

Discussion

This section seeks to highlight the differences between the tailored study identification protocol and the Cochrane study identification protocol as they relate to (i) the effectiveness of study identification, measured here by the number of studies identified and the number of studies identified uniquely, and (ii) the differences in the value of the studies, measured here by differences in study quality and the contribution to the synthesis of the studies identified. We focus on the primary study identification methods of the Cochrane study identification protocol (database searching) and the tailored study identification protocol (contacting organisations/web-searching), since these are ultimately the approaches by which the studies were uniquely identified in each case.

Effectiveness

Number of studies identified

The Cochrane study identification protocol identified 21,409 studies to screen compared to 453 studies identified by the tailored study identification protocol. Interpreting the difference between the tailored study identification protocol and the Cochrane study identification protocol in strictly numerical terms should be treated with caution since it risks overstating the efficiency of the tailored study identification protocol.

Prior to registering the review with The Cochrane Public Health Group, we had queried the utility of undertaking exhaustive and sensitive bibliographic database searches at the start of the review process. Researchers have found that even sensitive search strategies will not identify all studies in topics where a standardised or controlled terminology does not yet exist (Kwon et al., 2014, Golder and Loke, 2012), and key topic search terms for this review, nature or natural (for example), have multifarious application both as descriptors of place (i.e. adjectives) and also as definers of activity (i.e. adverbs). Defining a sufficiently sensitive literature search strategy, that produced a manageable number of search results to screen, represented a challenge, which was further compounded as standard techniques to improve efficiency in bibliographic database searches, such as the use of study design literature search filters, are not recommend in public health topics or reviews of conservation interventions (Bayliss and Beyer, 2015, Bayliss et al., 2014).

Contacting study authors and organisations as a primary method of study identification ameliorated some of these issues in the tailored study identification protocol. Previous studies have evaluated the effectiveness of contacting study authors to identify studies or study data (Gibson et al., 2006, Hetherington et al., 1989, McManus et al., 1998, Selph et al., 2014) but they have focused on the effectiveness of contact to identify data (as supported by our case study). We identified a further advantage: contacting study authors or organisations allowed us to explain our research question and inclusion criteria through conversation, circumventing the ambiguity of the search terms used in bibliographic database searching. Database hosts do not presently permit semantic searching, meaning that most search terms (indexing terms aside) do not differentiate retrieval based on meaning. Contacting relevant authors and organisations involved in the types of interventions under review allowed us to explain our research questions and this explains the lower number of studies identified. A positive side effect was to develop awareness and interest in our review from practitioners and policy makers.

In terms of effectively identifying studies and study data, our findings accord with other study authors who also report that contacting authors and experts will identify studies missed by bibliographic database searching (Haddaway and Bayliss, 2015, Westphal et al., 2014). Improved effectiveness should not, however, be confused with improved efficiency. We are comparing the searches retrospectively, and did not record the time taken to identify included studies using the Cochrane study identification protocol or the tailored study identification protocol at the time of the original review, but we conservatively estimate that the process of searching and screening in the Cochrane study identification protocol, and contacting organisations and web searching in the tailored study identification protocol, were approximately equal. The process of contacting organisations and web-searching is time intensive (Adams et al., 2016, Selph et al., 2014) with accompanying problems of data management and replicability (Adams et al., 2016). Bibliographic databases, almost without exception in this review, have export facilities to bibliographic management tools, whereas managing and de-duplicating studies identified through organisation contact and web-searching required manually entering study data into a bibliographic tool for screening (Stansfield et al., 2016).

Number of studies identified uniquely

After screening, the Cochrane study identification protocol identified two studies uniquely (Burls, 2007, Gooch, 2005) and the tailored study identification protocol identified nine studies uniquely: four using quantitative study designs (Brooker and Brooker, 2008a, Brooker and Brooker, 2008b, Eastaugh et al., 2010, Small Woods, 2011), two qualitative studies (Christie, 2004, Halpenny and Caissie, 2003) and three mixed-methods studies (BTCV, 2010, Wilson, 2009, Yerrell, 2008).

All studies using quantitative designs were identified by the tailored study identification protocol, whereas two qualitative studies were missed by the tailored study identification protocol. Understanding why the two qualitative studies were missed by the tailored study identification protocol would be almost impossible to unpick, since it would require re-contacting 288 organisations to ask them why they did not recommend those two studies. We explore the value of these two missed studies to the synthesis, and therefore develop our understanding of the significance of missing these studies in the tailored study identification protocol below, under study value.

Methodologically, the process of screening the 21,409 studies (31 days work at 7hrs a day/ screening at a rate of 100 studies per hour) identified in the Cochrane study identification protocol in order to identify two unique studies validates our initial concern that this topic was not necessarily suitable – or perhaps the topic area was not yet mature enough – for relying upon the application of sensitive, systematic bibliographic database searching. Researchers have previously questioned the utility of extensive online searches when compared with contacting organisations likely to collect review-relevant data (Haddaway and Bayliss, 2015, Bayliss and Beyer, 2015), and our findings in this study would support the usefulness of contacting organisations. Indeed, it could be worth questioning the practicable need for exhaustive bibliographic database searches in topics with a disparate evidence base (such as public health topics), or fields of research new to the techniques of systematic review, since the comprehensive identification of studies is often not an attainable goal.

It should be noted that the tailored study identification protocol did not directly compete against use of bibliographic database searches. As shown in figure one, we proposed to undertake bibliographic database searches as a supplement (i.e. adjunct), rather than as a primary method of study identification. We intended to use focused bibliographic database searches (Hausner et al., 2012), informed by our earlier grey literature searches. These

searches were not ultimately required, since we used the bibliographic database searches of the Cochrane study identification protocol as a surrogate.

Changing the chronological order of study identification methods from the Cochrane study identification protocol to the tailored study identification protocol may initially appear to be superficial but what we really seek to alter is the allocation of searching effort. This study confirms the value of aligning the primary method of study identification to where studies are most likely to be identified. In this case, the belief of our expert panel, that grey literature studies would be important to this review, meant we prioritised identification and searching effort for such studies over formally published studies indexed in bibliographic databases. The idea that the chronological order of study identification methods, led by a primary method of study identification, reflects the likely location of studies and affects the distribution of searching effort is not without precedent, since it forms the basis of the Cochrane study identification protocol. In the Cochrane study identification protocol, the information need (typically for studies reporting RCTs) is matched to a corresponding process of study identification. Generically, the process of study identification, as conducted by an expert searcher, can be perceived as starting from the methods most likely to identify relevant studies (and most likely to identify the most studies) to methods least likely to identify studies. Searching end-to-end of this methodological process seeks to address the risk of publication bias, since even those studies that are more difficult to identify are still sought, although in reality the time spent searching, using each individual search method, is often different and decreases after the primary method is undertaken. Hartling et al explore the possibility of prioritising which databases to search in systematic reviews (Hartling et al., 2016) but we believe this study is the first to prioritise and allocate search methods, in particular, supplementary search methods, in a review.

Studies have demonstrated (Helmer et al., 2001) or explored (Greenhalgh and Peacock, 2005) the use of supplementary search methods but our findings would suggest that categorising study identification methods as primary or supplementary is unhelpful, since no guidance exists on which search methods should be used for different review needs (Westphal et al., 2014). Our findings suggest that matching methods of study identification to the evidence base proved valuable in this case study and this approach may hold value not only for similar topics but also for other topic areas with a disparate evidence base.

Study value

Studies that evaluate search effectiveness commonly interpret effectiveness as the identification of studies missed when measured against a comparator or alternative search approach (Booth, 2010). Additional studies identified by alternative search methods can provide valuable information to researchers but the perceived value of those newly identified studies is seldom established and is difficult to measure accurately (Kwon et al., 2014).

Study quality

Quantitative

As Table 1 illustrates, all identified quantitative studies, both formally published (identified by the Cochrane study identification protocol and tailored study identification protocol) and grey literature studies (tailored study identification protocol only) were appraised as being of weak study quality in our Cochrane review. There is no perceivable improvement in study quality between the grey and published studies identified by the tailored study identification protocol, a finding that is consistent with other studies (Egger et al., 2003).

Qualitative

Conversely, there was a difference in study quality between the tailored study identification protocol and the Cochrane study identification protocol (Table 1). Three grey literature studies identified only by the tailored study identification protocol (Christie, 2004, Halpenny and Caissie, 2003, Wilson, 2009) scored one category higher on the Wallace criterion than the two published studies identified only in the Cochrane study identification protocol (Burls, 2007, Gooch, 2005). It is possible that the unpublished nature of the grey literature, with no limitation on the use of tables or words count, meant that greater detail was provided on the methods and results than would be possible in a journal article study. We interpret this idea cautiously, since the number of studies concerned is limited, and there is no wider empirical evidence to aid interpretation of this finding. Moreover, it does not follow that because greater detail is provided on the methods and results, that the study is generally of better quality.

Contribution to the synthesis

Quantitative

Comprehensive study identification is an important part of evaluating intervention effectiveness as it is linked to producing a reliable estimate of intervention effectiveness (Egger et al., 2003). The fact that the Cochrane study identification protocol would have missed nine studies (four quantitative and three mixed-methods) evaluating the effectiveness of environmental enhancement and conservation activity interventions is an important finding when considering the contribution of the tailored study identification protocol to the synthesis of effectiveness studies in this field. It highlights the importance of so-called 'supplementary search methods', perhaps suggesting that they are in fact complementary (possibly primary) methods of study identification.

Qualitative

With the qualitative studies, we found that two studies made no significant contribution to the synthesis and we therefore question the value of these studies in the synthesis and the impact of identifying them. We conclude that, had these studies been missed in study identification, the impact on the synthesis would have been negligible.

The study by Burls and the study by Gooch were uniquely identified by the Cochrane study identification protocol and after screening a significant number of non-relevant studies. We initially questioned the need for, and utility of, comprehensive bibliographic database searches in this review. Whilst this perception is only now clear through retrospective analysis, the research waste in searching, screening and ordering full-text in the Cochrane study identification protocol is potentially troubling, especially since we questioned the utility of comprehensive searching at the outset. We lacked the metric to test or demonstrate our concerns beyond suspicion. A metric to formatively test the effectiveness of study identification would be a valuable contribution to the process of systematic review.

Our findings in this case study raises further questions as to whether it is possible to conduct truly "comprehensive" searches for reviews (or topics) in which the evidence is widely dispersed across both academic databases and the 'grey literature,' and it highlights the need for so-called supplementary study identification methods (Helmer et al., 2001). Given the specific findings from the qualitative studies, this argument could be extended to reviews of qualitative studies: specifically that comprehensive study identification is unlikely to prove an attainable goal in most cases (Lorenc et al., 2012).

In retrospectively analysing both study identification protocols, we feel that the time invested in scoping, working with the PRG, and the make-up of our research team and team discussion, was of great benefit in developing the tailored study identification protocol. Linking the methods and process of study identification to study quality, or contribution of studies to

synthesis, could help researchers better understand the value of investing in the process of study identification or selecting more appropriate study identification methods. Matching methods of study identification to studies, and potentially working out when (or how) not to search, could yield benefits in the efficiency of study identification in systematic reviews.

Study limitations

The use of a case study research design to report this study means that the findings should be interpreted with caution since they relate to a single case study.

A limitation of this study is that time taken to undertake each individual search method was not recorded. This limits any interpretation as to the efficiency of the tailored study identification protocol and Cochrane study identification protocol. Recording time taken to search more generally would develop the evidence on the effectiveness and efficiency of searching in systematic reviews.

The quality of the studies identified and included in our Cochrane review was variable, which prohibits not only the interpretation of results and the conclusions that can be drawn from The Cochrane Review but also, it inhibits our ability to interpret the contribution of the study identification and to make links to study value. Better quality studies would aid interpretation and discussion.

Our use of CERQual to explore the contribution of the qualitative studies might be considered a limitation since its discriminant validity is yet to be established. Nevertheless, the use of CERQual in a supportive capacity reduces the dependence of the results on this specific tool.

Conclusions

In this study, we sought to link the idea of search effectiveness to study value. We retrospectively found that, in the case of a mixed methods review of a topic that crossed environmental and public health boundaries, extensive bibliographic database searching was of no value in terms of contribution to synthesis but that grey literature searching was valuable and identified studies that made unique contributions to both the quantitative and qualitative synthesis.

What we demonstrate in this case study is that the sequential order of study identification methods can be altered from a conventional study identification protocol. This, in effect, gives study identification methods different weighting depending upon how much effort and time is invested in them relative to the anticipated value. In the tailored study identification protocol, our primary methods of study identification were grey literature searching and contacting experts, which we demonstrate contributed valuable studies and study data. We valued bibliographic database searching as lower priority, so aimed to treat it as a supplementary study identification method, which, by comparing with the Cochrane study identification protocol, was valid.

ADAMS, J., HILLIER-BROWN, F. C., MOORE, H. J., LAKE, A. A., ARAUJO-SOARES, V., WHITE, M. & SUMMERBELL, C. 2016. Searching and synthesising 'grey literature' and 'grey information' in public health: critical reflections on three case studies. *Systematic Reviews,* 5, 164.

ARMIJO-OLIVO, S., STILES, C. R., HAGEN, N. A., BIONDO, P. D. & CUMMINGS, G. G. 2012. Assessment of study quality for systematic reviews: a comparison of the Cochrane Collaboration Risk of Bias Tool and the Effective Public Health Practice Project Quality Assessment Tool: methodological research. *J Eval Clin Pract,* 18, 12-8.

ARMSTRONG R, W. E., DOYLE J 2011. Chapter 21: Reviews in health promotion and public health. *In:* HIGGINS JPT, G. S. (ed.) *Cochrane Handbook for Systematic Reviews of Interventions* The Cochrane Collaboration.

BARTON, J. 2009. *The effects of green exercise on psychological health and well-being*. PhD, University of Essex.

BAYLISS, H. R. & BEYER, F. R. 2015. Information retrieval for ecological syntheses. *Research Synthesis Methods,* 6, 136-148.

BAYLISS, S. E., DAVENPORT, C. F. & PENNANT, M. E. 2014. Where and how to search for information on the effectiveness of public health interventions--a case study for prevention of cardiovascular disease. *Health Info Libr J,* 31, 303-13.

BEAHLER, C. C., SUNDHEIM, J. J. & TRAPP, N. I. 2000. Information retrieval in systematic reviews: challenges in the public health arena. *Am J Prev Med,* 18, 6-10.

BILOTTA, G. S., MILNER, A. M. & BOYD, I. 2014a. On the use of systematic reviews to inform environmental policies. *Environmental Science & Policy,* 42, 67-77.

BILOTTA, G. S., MILNER, A. M. & BOYD, I. L. 2014b. Quality assessment tools for evidence from environmental science. *Environmental Evidence,* 3, 14.

BIRCH, M. 2005. Cultivating wildness : three conservation volunteers' experiences of participation in the Green Gym scheme. *British Journal of Occupational Therapy,* 68, 244-252.

BOOTH, A. 2010. How much searching is enough? Comprehensive versus optimal retrieval for technology assessments. *International Journal of Technology Assessment in Health Care,* 26, 431-435.

BROOKER, J. & BROOKER, M. 2008a. Comparative exercise values of green gym and conventional gym: a personal evaluation. *WALLINGFORD GREEN GYM: exercise evaluation.*

BROOKER, J. & BROOKER, M. 2008b. Comparative heart rates following green gym, other outdoor exercise and conventional gym: a personal evaluation. *WALLINGFORD GREEN GYM: post-exercise evaluation.*

BTCV 2010. Wellbeing Comes Naturally: Year One Report. Doncaster: BTCV.

BURLS, A. 2007. People and green spaces: promoting public health and mental health well-being though ecotherapy. *Journal of Public Mental Health,* 6, 24-39. 2007.

CARTER, C. 2008. Offenders and Nature schemes: using conservation and forest management in rehabilitation. *Research Summary*. Farnham: Forest Research.

CENTRE FOR REVIEWS AND DISSEMINATION & (CRD) 2009. Systematic Reviews: CRD's guidance for undertaking reviews in health care

CHRISTIE, J. 2004. Volunteer attitudes and motivations: research findings and their application for sustainable community involvement programs in natural resource management. *Effective Sustainability Education: What Works? Why? Where Next? Linking Research and Practice*. 18-20 February 2004, Sydney, Australia.

DICKS, L. V., HODGE, I., RANDALL, N. P., SCHARLEMANN, J. P. W., SIRIWARDENA, G. M., SMITH, H. G., SMITH, R. K. & SUTHERLAND, W. J. 2014. A Transparent Process for "Evidence-Informed" Policy Making. *Conservation Letters,* 7, 119-125.

EASTAUGH, K., TUDGE, K. & LAWES, K. 2010. Wye Wood Evaluation 2006-2009. Telford: Small Woods Association.

EFFECTIVE PUBLIC HEALTH PRACTICE PROJECT. 2009. *Effective Public Health Practice Project: Quality Assessment Tool for Quantitative Studies* [Online]. Available: http://www.ephpp.ca/tools.html [Accessed 2017].

EGGER, M., JUNI, P., BARTLETT, C., HOLENSTEIN, F. & STERNE, J. 2003. How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. *Health Technol Assess,* 7, 1-76.

FAZEY, I., SALISBURY, J. G., LINDENMAYER, D. B., MAINDONALD, J. & DOUGLAS, R. 2004. Can methods applied in medicine be used to summarize and disseminate conservation research? *Environmental Conservation,* 31, 190-198.

GARSIDE, R. 2014. Should we appraise the quality of qualitative research reports for systematic reviews, and if so, how? *Innovation: The European Journal of Social Science Research,* 27, 67-79.

GIBSON, C. A., BAILEY, B. W., CARPER, M. J., LECHEMINANT, J. D., KIRK, E. P., HUANG, G., DUBOSE, K. D. & DONNELLY, J. E. 2006. Author contacts for retrieval of data for a meta-analysis on exercise and diet restriction. *International Journal of Technology Assessment in Health Care,* 22, 267-270.

GODIN, K., STAPLETON, J., KIRKPATRICK, S. I., HANNING, R. M. & LEATHERDALE, S. T. 2015. Applying systematic review search methods to the grey literature: a case study examining guidelines for school-based breakfast programs in Canada. *Systematic Reviews,* 4, 138.

GOLDER, S. & LOKE, Y. K. 2012. The contribution of different information sources for adverse effects data. *Int J Technol Assess Health Care,* 28, 133-7.

GOOCH, M. 2005. Voices of the volunteers: an exploration of the experiences of catchment volunteers in coastal Queensland, Australia. *Local Environment.Feb 2005 10 2005.*

GRAYSON L, G. A. 2003. A Difficult Business: Finding the Evidence for Social Science Reviews. *ESRC UK Centre for Evidence Based Policy and Practice.*

HADDAWAY, N. R. & BAYLISS, H. R. 2015. Shades of grey: Two forms of grey literature important for reviews in conservation. *Biological Conservation,* 191, 827-829.

HALPENNY, E. A. & CAISSIE, L. T. 2003. Volunteering on nature conservation projects: volunteer experience, attitudes and values. *Tourism Recreation Research,* 28, 25-33.

HARTLING, L., FEATHERSTONE, R., NUSPL, M., SHAVE, K., DRYDEN, D. M. & VANDERMEER, B. 2016. The contribution of databases to the results of systematic reviews: a cross-sectional study. *BMC Medical Research Methodology,* 16, 127.

HAUSNER, E., WAFFENSCHMIDT, S., KAISER, T. & SIMON, M. 2012. Routine development of objectively derived search strategies. *Systematic Reviews,* 1, 19.

HELMER, D., SAVOIE, I., GREEN, C. & KAZANJIAN, A. 2001. Evidence-based practice: extending the search to find material for the systematic review. *Bulletin of the Medical Library Association,* 89, 346-352.

HETHERINGTON, J., DICKERSIN, K., CHALMERS, I. & MEINERT, C. L. 1989. Retrospective and prospective identification of unpublished controlled trials: lessons from a survey of obstetricians and pediatricians. *Pediatrics,* 84, 374-80.

HUSK, KERRYN, LOVELL, REBECCA, COOPER, CHRIS, GARSIDE & RUTH. 2013. Participation in environmental enhancement and conservation activities for health and well-being in adults. *Cochrane Database of Systematic Reviews* [Online]. Available: http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD010351/abstract.

HUSK, K., LOVELL, R., COOPER, C., STAHL-TIMMINS, W. & GARSIDE, R. 2016. Participation in environmental enhancement and conservation activities for health and well-being in adults: a review of quantitative and qualitative evidence. *Cochrane Database of Systematic Reviews* [Online]. Available: http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD010351.pub2/abstract.

KAREIVA, P., MARVIER, M., WEST, S. & HORNISHER, J. 2002. Slow-moving journals hinder conservation efforts. *Nature,* 420, 15-15.

KWON, Y., POWELSON, S. E., WONG, H., GHALI, W. A. & CONLY, J. M. 2014. An assessment of the efficacy of searching in biomedical databases beyond MEDLINE in identifying studies for a systematic review on ward closures as an infection control intervention to control outbreaks. *Syst Rev,* 3, 135.

LEFEBVRE, C., MANHEIMER, E. & GLANVILLE, J. (ed.) 2011. *Chapter 6: Searching for studies*: The Cochrane Collaboration.

LEWIN, S., GLENTON, C., MUNTHE-KAAS, H., CARLSEN, B., COLVIN, C. J., GULMEZOGLU, M., NOYES, J., BOOTH, A., GARSIDE, R. & RASHIDIAN, A. 2015. Using qualitative evidence in decision making for health and social interventions: an approach to assess confidence in findings from qualitative evidence syntheses (GRADE-CERQual). *PLoS Med,* 12, e1001895.

LORENC, T., PEARSON, M., JAMAL, F., COOPER, C. & GARSIDE, R. 2012. The role of systematic reviews of qualitative evidence in evaluating interventions: a case study. *Research Synthesis Methods,* 3, 1-10.

LOVELL, R., HUSK, K., COOPER, C., STAHL-TIMMINS, W. & GARSIDE, R. 2015. Understanding how environmental enhancement and conservation activities may benefit health and wellbeing: a systematic review. *BMC Public Health,* 15, 864.

MAHOOD, Q., VAN EERD, D. & IRVIN, E. 2014. Searching for grey literature for systematic reviews: challenges and benefits. *Research Synthesis Methods,* 5, 221-234.

MCMANUS, R. J., WILSON, S., DELANEY, B. C., FITZMAURICE, D. A., HYDE, C. J., TOBIAS, R. S., JOWETT, S. & HOBBS, F. D. R. 1998. Review of the usefulness of contacting other experts when conducting a literature search for systematic reviews. *BMJ,* 317, 1562-1563.

MOORE, M., TOWNSEND, M. & OLDROYD, J. 2006. Linking Human and Ecosystem Health: The Benefits of Community Involvement in Conservation Groups. *EcoHealth,* 3, 255-261.

O'BRIEN, L., BURLS, A., TOWNSEND, M. & EBDEN, M. 2010. Volunteering in nature as a way of enabling people to re-integrate into society. *Perspectives in Public Health.*

O'BRIEN, L., TOWNSEND, M. & EBDEN, M. 2008. 'I like to think when I'm gone I will have left this a better place' Environmental volunteering: motivations, barriers and benefits. Scottish Forestry Trust and Forestry Commission.

PAPAIOANNOU, D., SUTTON, A., CARROLL, C., BOOTH, A. & WONG, R. 2010. Literature searching for social science systematic reviews: consideration of a range of search techniques. *Health Info Libr J,* 27, 114-22.

PILLEMER, K. 2010. Environmental volunteering and health outcomes over a 20-year period. *Gerontologist,* 50, 594-602. 2010.

PULLIN, A. S. & KNIGHT, T. M. 2001. Effectiveness in Conservation Practice: Pointers from Medicine and Public Health

Efectividad de la Conservación Práctica: Indicadores de Medicina y Salud Pública. *Conservation Biology,* 15, 50-54.

REYNOLDS, V. 1999. The Green Gym Evaluation of a pilot project in Sonning Common, Oxfordshire. Oxford Centre for Health Care Research and Development (OCHRAD); Oxford Brookes.

SELPH, S. S., GINSBURG, A. D. & CHOU, R. 2014. Impact of contacting study authors to obtain additional data for systematic reviews: diagnostic accuracy studies for hepatic fibrosis. *Systematic Reviews,* 3, 107.

SMALL WOODS, A. 2011. Amazon Woman Hereford SF36 Analysis. Telford: Small Woods Association.

STANSFIELD, C., DICKSON, K. & BANGPAN, M. 2016. Exploring issues in the conduct of website searching and other online sources for systematic reviews: how can we be systematic? *Systematic Reviews,* 5, 191.

STERNE JAC, EGGER M & MOHER D (eds.) 2011. *Chapter 10: Addressing reporting biases*: The Cochrane Collaboration.

STEWART, G. B., COLES, C. F. & PULLIN, A. S. 2005. Applying evidence-based practice in conservation management: Lessons from the first systematic review and dissemination projects. *Biological Conservation,* 126, 270-278.

THOMAS, J. & HARDEN, A. 2008. Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology,* 8, 45.

TOWNSEND, M. 2006. Feel blue? Touch green! Participation in forest/woodland management as a treatment for depression. *Urban Forestry & Urban Greening; 2006.5: 3, 111-120.30 ref*.

TOWNSEND, M. & MARSH, R. 2004. Exploration of the Health and Well-being Benefits of Membership of Truganina Explosives Reserve Preservation Society. Burwood, Australia: School of Health and Social Development, Deakin University.

TOWNSEND, M. & MOORE, M. 2005. Research into the health, wellbeing & social capital benefits of community involvement in the management of land for conservation : final report. Geelong, Vic.: Deakin University, Trust for Nature.

WALLACE, A., CROUCHER, K., QUILGARS, D. & BALDWIN, S. 2004. Meeting the challenge: developing systematic reviewing in social policy. *Policy & Politics,* 32, 455-470.

WESTPHAL, A., KRISTON, L., HOLZEL, L. P., HARTER, M. & VON WOLFF, A. 2014. Efficiency and contribution of strategies for finding randomized controlled trials: a case study from a systematic review on therapeutic interventions of chronic depression. *J Public Health Res,* 3, 177.

WILSON, N. 2009. Branching Out. Greenspace and conservation on referral. Edinburgh: Forestry Commission Scotland, NHSGGC, Glsgow Centre for Population Health, Glasgow Clyde Valley Green Newtork Partnership.

YERRELL, P. 2008. National Evaluation of BTCV's Green Gym.

Table 1: Study Quality

Study	Study Type	Identification Method	EPHPP	Wallace
Brooker and Brooker 2008*	Quantitative	TSIP	Weak	
Brooker and Brooker 2008*	Quantitative	TSIP	Weak	
Eastaugh 2010	Quantitative	TSIP	Weak	
Small Woods 2011a	Quantitative	TSIP	Weak	
Barton 2009	Quantitative	CSIP + TSIP	Weak	
Pillemer 2010	Quantitative	CSIP + TSIP	Weak	
Reynolds 1999a	Quantitative	CSIP + TSIP	Weak	
Townsend 2005	Quantitative	CSIP + TSIP	Weak	
Christie 2004	Qualitative	TSIP		Good
Halpenny and Cassie 2003	Qualitative	TSIP		Good
Burls 2007	Qualitative	CSIP		Moderate
Gooch 2005	Qualitative	CSIP		Moderate
Birch 2005	Qualitative	CSIP + TSIP		Moderate
Carter 2008	Qualitative	CSIP + TSIP		Moderate
O'Brien 2010a	Qualitative	CSIP + TSIP		Good
Townsend 2006	Qualitative	CSIP + TSIP		Moderate
Townsend and Marsh 2004	Qualitative	Citation chase		Moderate
BTCV 2010	Mixed Methods	TSIP	Weak	Moderate
Wilson 2009	Mixed Methods	TSIP	Weak	Good
Yerrell 2008	Mixed Methods	TSIP	Weak	
O'Brien 2008a	Mixed Methods	CSIP + TSIP	Weak	Good

* studies were included in the review but excluded from the synthesis due to poor study quality. Key: TSIP = tailored study identification protocol and CSIP = Cochrane study identification protocol.

Table 2: *Quantitative results*

Study	Identification Method	Mental and Emotional Wellbeing			HRQoL			Physical Activity Measures		
		Reported	Tool	Outcome	Reported	Tool	Outcome	Reported	Tool	Outcome
Barton 2009	CSIP + TSIP	✓	RSES + PMSS	No change	x			x		
O'Brien 2008a	CSIP + TSIP	✓	ESS	Significant improvement	x			x		
Pillemer 2010	CSIP + TSIP	✓	NR	Reduction	✓	Retrospective comparison	Improvement with volunteers	✓	Unique to study	PA sig. associated with volunteers
Reynolds 1999a	CSIP + TSIP	x			✓	SF-36	Improvements*	x		
Townsend 2005	CSIP + TSIP	✓	NR	Some differences	✓	Likert scale	Some improvements	x		
BTCV 2010	TSIP	x			✓	SF-12	Little/no change	x		
Eastaugh 2010	TSIP	x			✓	SF-36	Little/no change	x		
Small Woods 2011a	TSIP	x			✓	SF-36	Improvements*	x		
Wilson 2009	TSIP	✓	WEMWBS	Increased or no change	✓	SF-12	Little/no change	✓	SPAQ	Increased PA
Yerrell 2008	TSIP	x			✓	PCS/MCS-12	Improvements	x		

Key: Emotional State Scale (ESS); Rosenberg self-esteem scale (RSES); Profile of Mood States scale (PMSS); physical activity (PA); Warwick-Edinburgh Mental Well-being Scale (WEMWBS); Scottish Physical Activity Questionnaire (SPAQ). CSIP = Cochrane study identification protocol and TSIP = tailored study identification protocol.
Notes: *very small sample sizes so robustness of results is questionable

Table 3: *Presence of qualitative themes in each study*

Author	Identification Method	Personal Achievement	Personal / Social Identify	Developing Knowledge	Benefits of place	Social Contact	Physical Activity	Spirituality	Psychological benefits	Risks/ negatives
Townsend & Marsh 2004*	Citation chase	✓	X	✓	✓	✓	✓	X	✓	X
		✓	X	✓	X	✓	✓	X	✓	X
Burls 2007	CSIP	✓	✓	✓	✓	✓	✓	✓	✓	X
Gooch 2005	CSIP	✓	✓	✓	✓	✓	X	X	✓	✓
Birch 2005	CSIP + TSIP	✓	X	X	✓	✓	✓	X	✓	X
Carter 2008	CSIP + TSIP	✓	✓	✓	✓	✓	✓	X	✓	X
O'Brien 2008a	CSIP + TSIP	✓	✓	✓	✓	✓	✓	✓	✓	X
O'Brien 2010a	CSIP + TSIP	✓	X	✓	✓	✓	X	✓	✓	X
Townsend 2006	CSIP + TSIP	✓	X	X	✓	✓	✓	X	✓	X
BTCV 2010*	TSIP	✓	X	✓	✓	✓	X	✓	✓	✓
		✓	X	✓	✓	✓	X	X	✓	✓
Christie 2004	TSIP	✓	✓	✓	✓	✓	X	✓	✓	✓
Halpenny & Cassie 2003	TSIP	✓	X	X	✓	✓	X	X	✓	X
Wilson 2009	TSIP	✓	✓	✓	✓	✓	X	X	X	✓

*there were two sub-groups for each of these citations
Key: TSIP = tailored study identification protocol and CSIP = Cochrane study identification protocol.

Table 4: *CERQual all studies included*

Review finding	studies contributing to the review finding	Assessment of methodological limitations	Assessment of relevance	Assessment of coherence	Assessment of adequacy	Overall CERQual assessment of confidence	Explanation of judgement
Physical activity	Seven studies. (Townsend & Marsh 2004[1*]; Burls 2007[2];	Minor methodological limitations	No concerns	No concerns	Minor concerns	Moderate confidence	This theme was graded as moderate confidence since there were minor

	Birch 2005[3]; Carter 2008[3]; O'Brien 2008a[3]; Townsend 2006[3]; Wilson 2009[4])	Two studies were rated as good (O'Brien 2008a[3]; Wilson 2009[4]) Five studies were rated as moderate (Townsend & Marsh 2004[1*]; Burls 2007[2]; Birch 2005[3]; Carter 2008[3]; Townsend 2006[3])					concerns on study quality and adequacy of data.
Personal achievement	Twelve studies (Townsend & Marsh 2004[1*]; Burls 2007[2]; Gooch 2005[2]; Birch 2005[3]; Carter 2008[3]; O'Brien 2008a[3]; O'Brien 2010a[3]; Townsend 2006[3]; BTCV 2010[4*]; Christie 2004[4]; Halpenny & Cassie 2003[4]; Wilson 2009[4])	No concerns Five studies rated as Good (Christie 2004[4]; Halpenny & Cassie 2003[4]; O'Brien 2008a[3]; O'Brien 2010a[3]; Wilson 2009[4]) Seven studies rated moderate (Townsend & Marsh 2004[1*]; Burls 2007[2]; Gooch 2005[2]; Birch 2005[3]; Carter 2008[3]; Townsend 2006[3]; BTCV 2010[4*])	No concerns	No concerns	No concerns	High confidence	This theme was graded as high confidence since there were no concerns in the four CERQual domains.

Personal/ Social Identity	Six studies (Carter 2008[3]; Christie 2004[4]; O'Brien 2008a[3]; Gooch 2005[2]; Wilson 2009[4]; Burls 2007[2])	No concerns Three studies were rated as good (Christie 2004[4]; O'Brien 2008a[3]; Wilson 2009[4]) Three studies were rated as moderate (Carter 2008[3]; Gooch 2005[2]; Burls 2007[2])	No concerns	No concerns	No concerns	High confidence	This theme was graded as high confidence since there were no concerns in the four CERQual domains.
Developing knowledge	Nine studies (Townsend & Marsh 2004[1*]; Burls 2007[2]; Gooch 2005[2]; Carter 2008[3]; O'Brien 2008a[3]; O'Brien 2010a[3]; BTCV 2010[4*]; Christie 2004[4]; Wilson 2009[4])	No concerns Four studies rated as good (Christie 2004[4]; O'Brien 2008a[3]; O'Brien 2010a[3]; Wilson 2009[4]) Five studies rated as moderate (Townsend & Marsh 2004[1*]; Burls 2007[2]; Gooch 2005[2]; Carter 2008[3]; BTCV 2010[4*])	No concerns	No concerns	No concerns	High confidence	This theme was graded as high confidence since there were no concerns in the four CERQual domains.
Benefits of place	Twelve studies (Townsend & Marsh 2004[1*]; Burls 2007[2];	No concerns Five studies rated as Good (Christie 2004[4]; Halpenny	No concerns	No concerns	No concerns	High confidence	This theme was graded as high confidence since there were no concerns in the

	Gooch 2005[2]; Birch 2005[3]; Carter 2008[3]; O'Brien 2008a[3]; O'Brien 2010a[3]; Townsend 2006[3]; BTCV 2010[4*]; Christie 2004[4]; Halpenny & Cassie 2003[4]; Wilson 2009[4])	& Cassie 2003[4]; O'Brien 2008a[3]; O'Brien 2010a[3]; Wilson 2009[4]) Seven studies rated moderate (Townsend & Marsh 2004[1*]; Burls 2007[2]; Gooch 2005[2]; Birch 2005[3]; Carter 2008[3]; Townsend 2006[3]; BTCV 2010[4*])					four CERQual domains.
Social contact	Twelve studies (Townsend & Marsh 2004[1*]; Burls 2007[2]; Gooch 2005[2]; Birch 2005[3]; Carter 2008[3]; O'Brien 2008a[3]; O'Brien 2010a[3]; Townsend 2006[3]; BTCV 2010[4*]; Christie 2004[4]; Halpenny & Cassie 2003[4]; Wilson 2009[4])	No concerns Five studies rated as Good (Christie 2004[4]; Halpenny & Cassie 2003[4]; O'Brien 2008a[3]; O'Brien 2010a[3]; Wilson 2009[4]) Seven studies rated moderate (Townsend & Marsh 2004[1*]; Burls 2007[2]; Gooch 2005[2]; Birch 2005[3]; Carter 2008[3];	No concerns	No concerns	No concerns	High confidence	This theme was graded as high confidence since there were no concerns in the four CERQual domains.

		Townsend 2006[3]; BTCV 2010[4*])					
Spirituality	Five studies (Burls 2007[2]; O'Brien 2008a[3]; O'Brien 2010a[3]; BTCV 2010[4*]; Christie 2004[4])	No concerns three studies were rated as good (O'Brien 2008a[3]; O'Brien 2010a[3]; Christie 2004[4]) two studies were rated as moderate (Burls 2007[2]; BTCV 2010[4*])	No concerns	No concerns	No concerns	High confidence	This theme was graded as high confidence since there were no concerns in the four CERQual domains.
Psychological benefits	Twelve studies (Townsend & Marsh 2004[1*]; Burls 2007[2]; Gooch 2005[2]; Birch 2005[3]; Carter 2008[3]; O'Brien 2008a[3]; O'Brien 2010a[3]; Townsend 2006[3]; BTCV 2010[4*]; Christie 2004[4]; Halpenny & Cassie 2003[4]; Wilson 2009[4])	No concerns Five studies rated as Good (Christie 2004[4]; Halpenny & Cassie 2003[4]; O'Brien 2008a[3]; O'Brien 2010a[3]; Wilson 2009[4]) Seven studies rated moderate (Townsend & Marsh 2004[1*]; Burls 2007[2]; Gooch 2005[2]; Birch 2005[3]; Carter 2008[3]; Townsend 2006[3]; BTCV 2010[4*])	No concerns	No concerns	No concerns	High confidence	This theme was graded as high confidence since there were no concerns in the four CERQual domains.

Risks and negative impacts	Four studies (Gooch 2005[2]; BTCV 2010[4*]; Christie 2004[4]; Wilson 2009[4])	No concerns Two studies were rated as good (Christie 2004[4]; Wilson 2009[4]) two studies were rated as moderate (Gooch 2005[2]; BTCV 2010[4*])	No concerns	No concerns	Minor concerns	Moderate confidence	This theme was graded as moderate confidence since there were minor concerns on the adequacy of data.

[1]Citation Chasing; [2] Cochrane study identification protocol; [3] Cochrane study identification protocol & Tailored study identification protocol, and; [4] Tailored study identification protocol. [*] there were two sub-groups for each of these citations.

Table 5: *CERQual Burls and Gooch removed*

Review finding	studies contributing to the review finding	Assessment of methodological limitations	Assessment of relevance	Assessment of coherence	Assessment of adequacy	Overall CERQual assessment of confidence	Explanation of judgement
Physical activity	Six studies. (Townsend & Marsh 2004[1*]; Birch 2005[3]; Carter 2008[3]; O'Brien 2008a[3]; Townsend 2006[3]; Wilson 2009[4])	Minor methodological limitations Two studies were rated as good (O'Brien 2008a[3]; Wilson 2009[4]) Four studies were rated as moderate (Townsend & Marsh 2004[1*];	No concerns	No concerns	No concerns	Moderate confidence	This theme was graded as moderate confidence since there were minor concerns on study quality. In this theme, Burls provides confirmatory validity alongside Birch for the same sub-theme. The loss

		Birch 2005[3]; Carter 2008[3]; Townsend 2006[3])					of Burls would therefore be insignificant.
Personal achievement	Ten studies (Townsend & Marsh 2004[1*]; Birch 2005[3]; Carter 2008[3]; O'Brien 2008a[3]; O'Brien 2010a[3]; Townsend 2006[3]; BTCV 2010[4*]; Christie 2004[4]; Halpenny & Cassie 2003[4]; Wilson 2009[4])	No concerns Five studies rated as Good (Christie 2004[4]; Halpenny & Cassie 2003[4]; O'Brien 2008a[3]; O'Brien 2010a[3]; Wilson 2009[4]) Five studies rated moderate (Townsend & Marsh 2004[1*]; Birch 2005[3]; Carter 2008[3]; Townsend 2006[3]; BTCV 2010[4*])	No concerns	No concerns	Minor concerns The loss of Burls removes some confirmatory richness as a participant quote would be lost. The study that defines the sub-theme of 'payback' (Christie 04) remains, so the underlying data is not lost. This theme is well supported by studies.	High confidence	This theme was graded as high confidence since the loss of confirmatory richness in the form of Burls, was considered a minor point in the identification of the theme and contribution to the synthesis. Similarly, Gooch provides confirmatory validity to a sub-theme already supported by other studies one of which (Christie 04) is of better methodological quality.
Personal/ Social Identity	Four studies (Carter 2008[3]; Christie 2004[4]; O'Brien 2008a[3]; Wilson 2009[4])	No concerns Three studies were rated as good (Christie 2004[4]; O'Brien 2008a[3]; Wilson 2009[4])	No concerns	No concerns	No concerns Neither the study by Burls or the study by Gooch provided either confirmatory richness or	High confidence	This theme was graded as high confidence since there were no concerns in the four CERQual domains.

		One study was rated as moderate (Carter 2008[3])			validity in this sub-theme. Moreover, neither study uniquely identified any subthemes.		The omission of both Burls and Gooch would not alter this theme.
Developing knowledge	Seven studies (Townsend & Marsh 2004[1*]; Carter 2008[3]; O'Brien 2008a[3]; O'Brien 2010a[3]; BTCV 2010[4*]; Christie 2004[4]; Wilson 2009[4])	No concerns Four studies rated as good (Christie 2004[4]; O'Brien 2008a[3]; O'Brien 2010a[3]; Wilson 2009[4]) Three studies rated as moderate (Townsend & Marsh 2004[1*]; Carter 2008[3]; BTCV 2010[4*])	No concerns	No concerns	No concerns The loss of Burls removes some validating richness. The loss of Gooch removes some confirmatory richness as a participant quote would be lost.	High confidence	This theme was graded as high confidence since the change in assessment of adequacy was felt to be minor resulting in no change to the synthesis.
Benefits of place	Ten studies (Townsend & Marsh 2004[1*]; Birch 2005[3]; Carter 2008[3]; O'Brien 2008a[3]; O'Brien 2010a[3]; Townsend 2006[3]; BTCV 2010[4*]; Christie 2004[4]; Halpenny &	No concerns Five studies rated as Good (Christie 2004[4]; Halpenny & Cassie 2003[4]; O'Brien 2008a[3]; O'Brien 2010a[3]; Wilson 2009[4]) Five studies rated moderate	No concerns	No concerns	No concerns The loss of Burls removes some confirmatory richness as the study is quoted three times. On each occasion, it is only to confirm or validate studies providing richer data.	High confidence	This theme was graded as high confidence since there were no concerns in the four CERQual domains. The loss of Burls was considered more important than the loss of Gooch but neither studies were

	Cassie 2003[4]; Wilson 2009[4]	(Townsend & Marsh 2004[1*]; Birch 2005[3]; Carter 2008[3]; Townsend 2006[3]; BTCV 2010[4*])					sufficiently valuable to alter the synthesis since neither study directly supported the identification of any sub-themes.
Social contact	Ten studies (Townsend & Marsh 2004[1*]; Birch 2005[3]; Carter 2008[3]; O'Brien 2008a[3]; O'Brien 2010a[3]; Townsend 2006[3]; BTCV 2010[4*]; Christie 2004[4]; Halpenny & Cassie 2003[4]; Wilson 2009[4])	No concerns Five studies rated as Good (Christie 2004[4]; Halpenny & Cassie 2003[4]; O'Brien 2008a[3]; O'Brien 2010a[3]; Wilson 2009[4]) Five studies rated moderate (Townsend & Marsh 2004[1*]; Birch 2005[3]; Carter 2008[3]; Townsend 2006[3]; BTCV 2010[4*])	No concerns	No concerns	Minor concerns Burls is not referenced in the synthesis. Gooch provides validating richness to one sub-theme.	High confidence	This theme was graded as high confidence. The minor concerns on adequacy are very minor concerns since neither study identified a sub-theme or provided confirmatory richness in the form of participant quotes.

Spirituality	Four studies (O'Brien 2008a[3]; O'Brien 2010a[3]; BTCV 2010[4*]; Christie 2004[4])	No concerns three studies were rated as good (O'Brien 2008a[3]; O'Brien 2010a[3]; Christie 2004[4]) one study was rated as moderate (BTCV 2010[4*])	No concerns	No concerns	No concerns The loss of Burls removes some validating richness but it is one of four studies cited in the identification of a sub-theme so the contribution of Burls is questionable.	High confidence	This theme was graded as high confidence since there were no concerns in the four CERQual domains.
Psychological benefits	Ten studies (Townsend & Marsh 2004[1*]; Burls 2007[2]; Gooch 2005[2]; Birch 2005[3]; Carter 2008[3]; O'Brien 2008a[3]; O'Brien 2010a[3]; Townsend 2006[3]; BTCV 2010[4*]; Christie 2004[4]; Halpenny & Cassie 2003[4]; Wilson 2009[4])	No concerns Five studies rated as Good (Christie 2004[4]; Halpenny & Cassie 2003[4]; O'Brien 2008a[3]; O'Brien 2010a[3]; Wilson 2009[4]) Five studies rated moderate (Townsend & Marsh 2004[1*]; Birch 2005[3]; Carter 2008[3]; Townsend 2006[3]; BTCV 2010[4*])	No concerns	No concerns	No concerns	High confidence	This theme was graded as high confidence since there were no concerns in the four CERQual domains.

Risks and negative impacts	Three studies (BTCV 2010[4*]; Christie 2004[4]; Wilson 2009[4])	No concerns Two studies were rated as good (Christie 2004[4]; Wilson 2009[4]) one study was rated as moderate (BTCV 2010[4*])	No concerns	No concerns	Minor concerns	moderate confidence	This theme was graded as moderate confidence since there were minor concerns on the adequacy of data.

[1]Citation Chasing; [2] Cochrane study identification protocol; [3] Cochrane study identification protocol & Tailored study identification protocol, and; [4] Tailored study identification protocol. [*] there were two sub-groups for each of these citations.

Table 6: *Christie and Halpenny & Cassie removed*

Review finding	studies contributing to the review finding	Assessment of methodological limitations	Assessment of relevance	Assessment of coherence	Assessment of adequacy	Overall CERQual assessment of confidence	Explanation of judgement
Physical activity	Six studies. (Townsend & Marsh 2004[1*]; Birch 2005[3]; Carter 2008[3]; O'Brien 2008a[3]; Townsend 2006[3]; Wilson 2009[4])	Minor methodological limitations Two studies were rated as good (O'Brien 2008a[3]; Wilson 2009[4])	No concerns	No concerns	No concerns	Moderate confidence	This theme was graded as moderate confidence since there were minor concerns on study quality. Christie and Halpenny and Cassie did not

		Four studies were rated as moderate (Townsend & Marsh 2004[1*]; Birch 2005[3]; Carter 2008[3]; Townsend 2006[3])					contribute to this theme so there are no changes to the CERQual judgement.
Personal achievement	Eight studies (Townsend & Marsh 2004[1*]; Birch 2005[3]; Carter 2008[3]; O'Brien 2008a[3]; O'Brien 2010a[3]; Townsend 2006[3]; BTCV 2010[4*]; 2003[4]; Wilson 2009[4])	Moderate concerns Three studies rated as Good (O'Brien 2008a[3]; O'Brien 2010a[3]; Wilson 2009[4]) Five studies rated moderate (Townsend & Marsh 2004[1*]; Birch 2005[3]; Carter 2008[3]; Townsend 2006[3]; BTCV 2010[4*])	No concerns	Minor concerns The loss of Christie represents the loss of relevant data to support and identify sub-themes. The loss of Christie therefore raises questions about the coherence of the sub-themes since Christie identifies sub-themes that are supported by other weaker studies.	Major concerns The loss of Christie represents the loss of relevant data and a key study. Sub-themes would have been missed.	Low confidence	This theme was graded as low confidence. The loss of Christie & Halpenny and Cassie represent the loss of two 'good' quality studies from this theme. The loss of Christie, specifically, represents the loss of what we consider a key study to this theme which, in terms of adequacy would mean two sub-themes would have been missed.
Personal/ Social Identity	Three studies (Carter 2008[3]; O'Brien 2008a[3]; Wilson 2009[4])	Moderate concerns Two studies were rated as good (O'Brien 2008a[3]; Wilson 2009[4])	No concerns	Moderate concerns The data on the sub-theme of identity being linked to the	Minor concerns In comparison to other themes, this theme was weakly supported by	Moderate confidence	This theme was graded as moderate confidence. The omission of Christie would

		One study was rated as moderate (Carter 2008[3])		impact in the environment was incoherent. Christie was the only 'good quality' study in the identification of this sub-theme and it provided data that contrasted with other studies.	study data. The loss of Christie as a key study raises concerns.		alter the understanding of this theme in the synthesis of studies.
Developing knowledge	Six studies (Townsend & Marsh 2004[1*]; Carter 2008[3]; O'Brien 2008a[3]; O'Brien 2010a[3]; BTCV 2010[4*]; Wilson 2009[4])	No concerns Three studies rated as good (O'Brien 2008a[3]; O'Brien 2010a[3]; Wilson 2009[4]) Three studies rated as moderate (Townsend & Marsh 2004[1*]; Carter 2008[3]; BTCV 2010[4*])	No concerns	No concerns	No concerns	High confidence	This theme was graded as high confidence.
Benefits of place	Eight studies (Townsend & Marsh 2004[1*]; Birch 2005[3]; Carter 2008[3]; O'Brien 2008a[3]; O'Brien 2010a[3];	Minor concerns Three studies rated as Good (O'Brien 2008a[3]; O'Brien 2010a[3]; Wilson 2009[4])	No concerns	No concerns	Minor concerns Removing Christie removes some validating richness through the loss of participant	Moderate confidence	This theme was graded as moderate confidence since there were minor concerns in the two CERQual domains.

	Townsend 2006[3]; BTCV 2010[4*]; Wilson 2009[4]	Five studies rated moderate (Townsend & Marsh 2004[1*]; Birch 2005[3]; Carter 2008[3]; Townsend 2006[3]; BTCV 2010[4*])			quotes to support sub-themes. Other, weaker, studies do provide data, however.		
Social contact	Eight studies (Townsend & Marsh 2004[1*]; Birch 2005[3]; Carter 2008[3]; O'Brien 2008a[3]; O'Brien 2010a[3]; Townsend 2006[3]; BTCV 2010[4*]; Wilson 2009[4])	Minor concerns Three studies rated as Good (O'Brien 2008a[3]; O'Brien 2010a[3]; Wilson 2009[4]) Five studies rated moderate (Townsend & Marsh 2004[1*]; Birch 2005[3]; Carter 2008[3]; Townsend 2006[3]; BTCV 2010[4*])	No concerns	No concerns	Minor concerns	Moderate confidence	This theme was graded as Moderate confidence

Spirituality	Three studies (O'Brien 2008a[3]; O'Brien 2010a[3]; BTCV 2010[4*])	No concerns two studies were rated as good (O'Brien 2008a[3]; O'Brien 2010a[3];) one study was rated as moderate (BTCV 2010[4*])	No concerns	No concerns	Major concerns The loss of Christie would prohibit the identification of one (out of two) sub themes.	Low confidence	This theme was graded as low confidence since there was major concerns on data adequacy.
Psychological benefits	Eight studies (Townsend & Marsh 2004[1*]; Burls 2007[2]; Gooch 2005[2]; Birch 2005[3]; Carter 2008[3]; O'Brien 2008a[3]; O'Brien 2010a[3]; Townsend 2006[3]; BTCV 2010[4*]; Wilson 2009[4])	No concerns Three studies rated as Good (O'Brien 2008a[3]; O'Brien 2010a[3]; Wilson 2009[4]) Five studies rated moderate (Townsend & Marsh 2004[1*]; Birch 2005[3]; Carter 2008[3]; Townsend 2006[3]; BTCV 2010[4*])	No concerns	No concerns	No concerns	High confidence	This theme was graded as high confidence since there were no concerns in the four CERQual domains.

Risks and negative impacts	Two studies (BTCV 2010[4*]; Wilson 2009[4])	No concerns One study was rated as good (Wilson 2009[4]) one study was rated as moderate (BTCV 2010[4*])	No concerns	No concerns	Minor concerns	moderate confidence	This theme was graded as moderate confidence since there were minor concerns on the adequacy of data.

[1]Citation Chasing; [2] Cochrane study identification protocol; [3] Cochrane study identification protocol & Tailored study identification protocol, and; [4] Tailored study identification protocol. [*] there were two sub-groups for each of these citations.

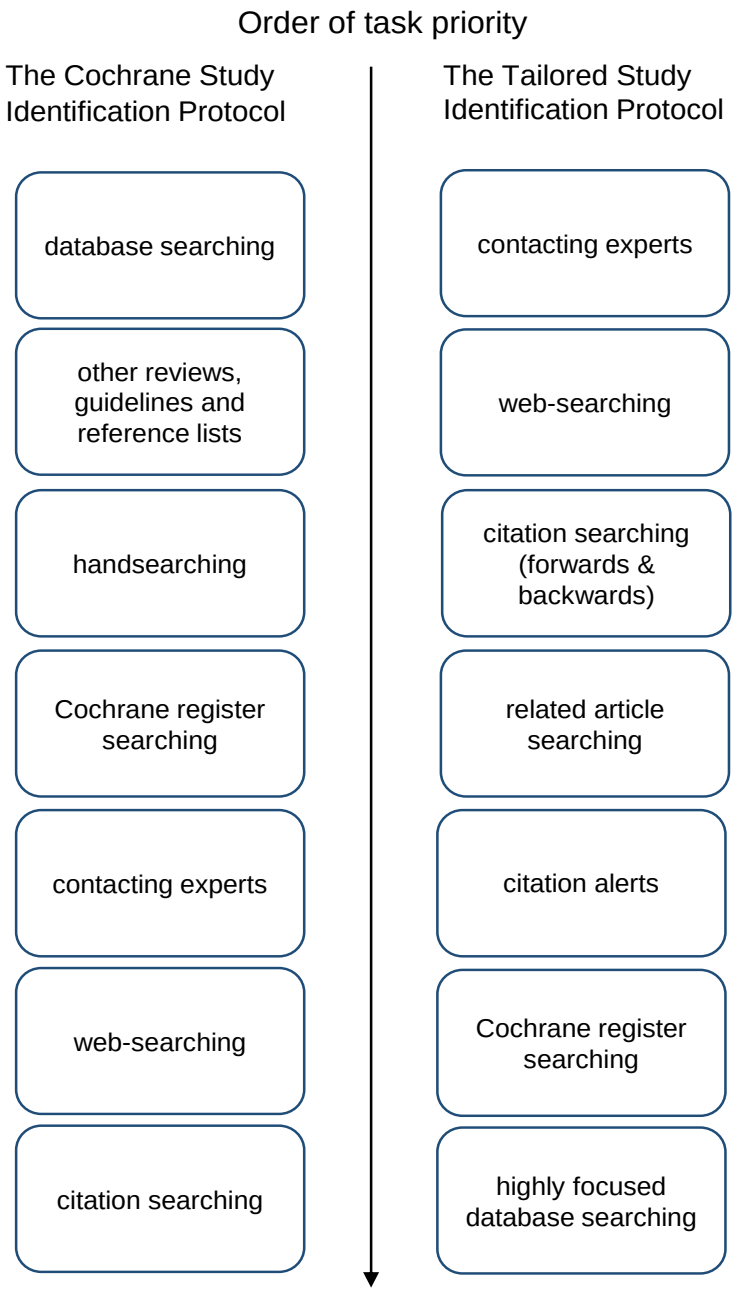
Order of task priority
The Cochrane Study Identification Protocol
The Tailored Study Identification Protocol
database searching
other reviews, guidelines and reference lists
handsearching
Cochrane register searching
contacting experts
web-searching
citation searching
contacting experts
web-searching
citation searching (forwards & backwards)
related article searching
citation alerts
Cochrane register searching
highly focused database searching

Figure one: *Schematic of Cochrane protocol and the Tailored protocol*, showing the primary and supplementary methods of study identification, and the chronological order and investment in study identification methods.

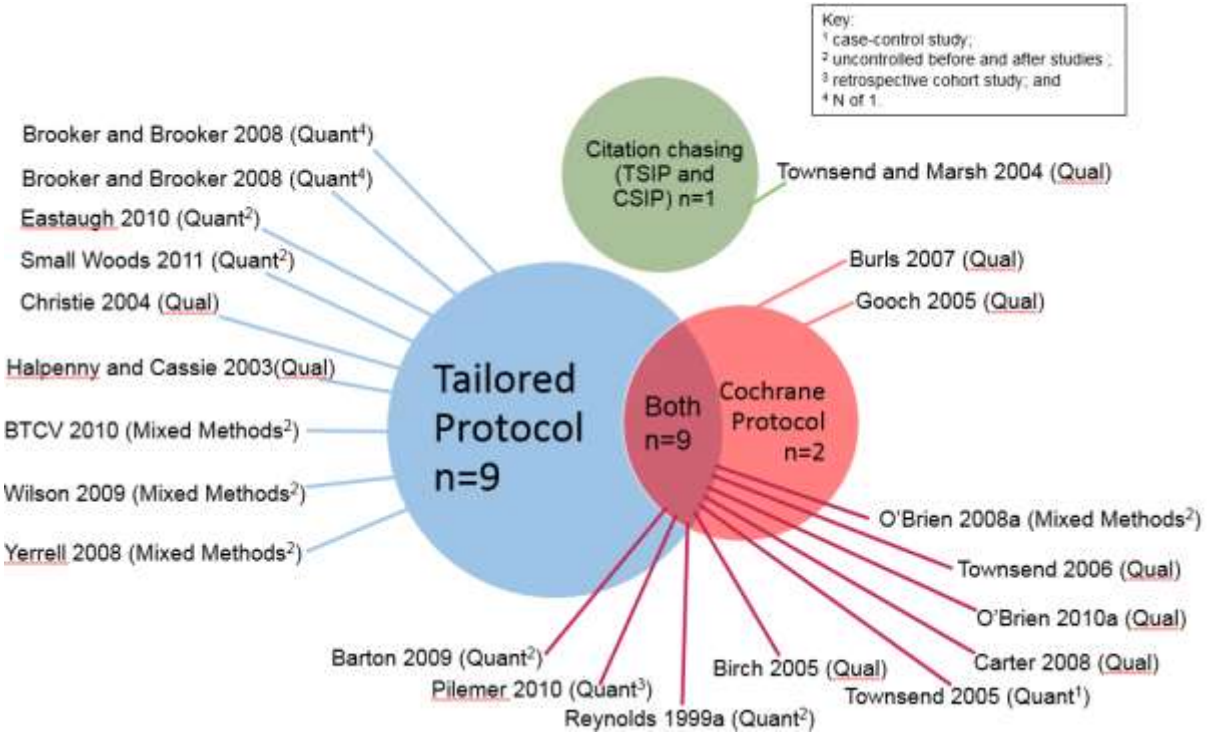

Figure two: *schematic of source of study identification. Key: TSIP = Tailored study identification protocol and CSIP = Cochrane study identification protocol.*

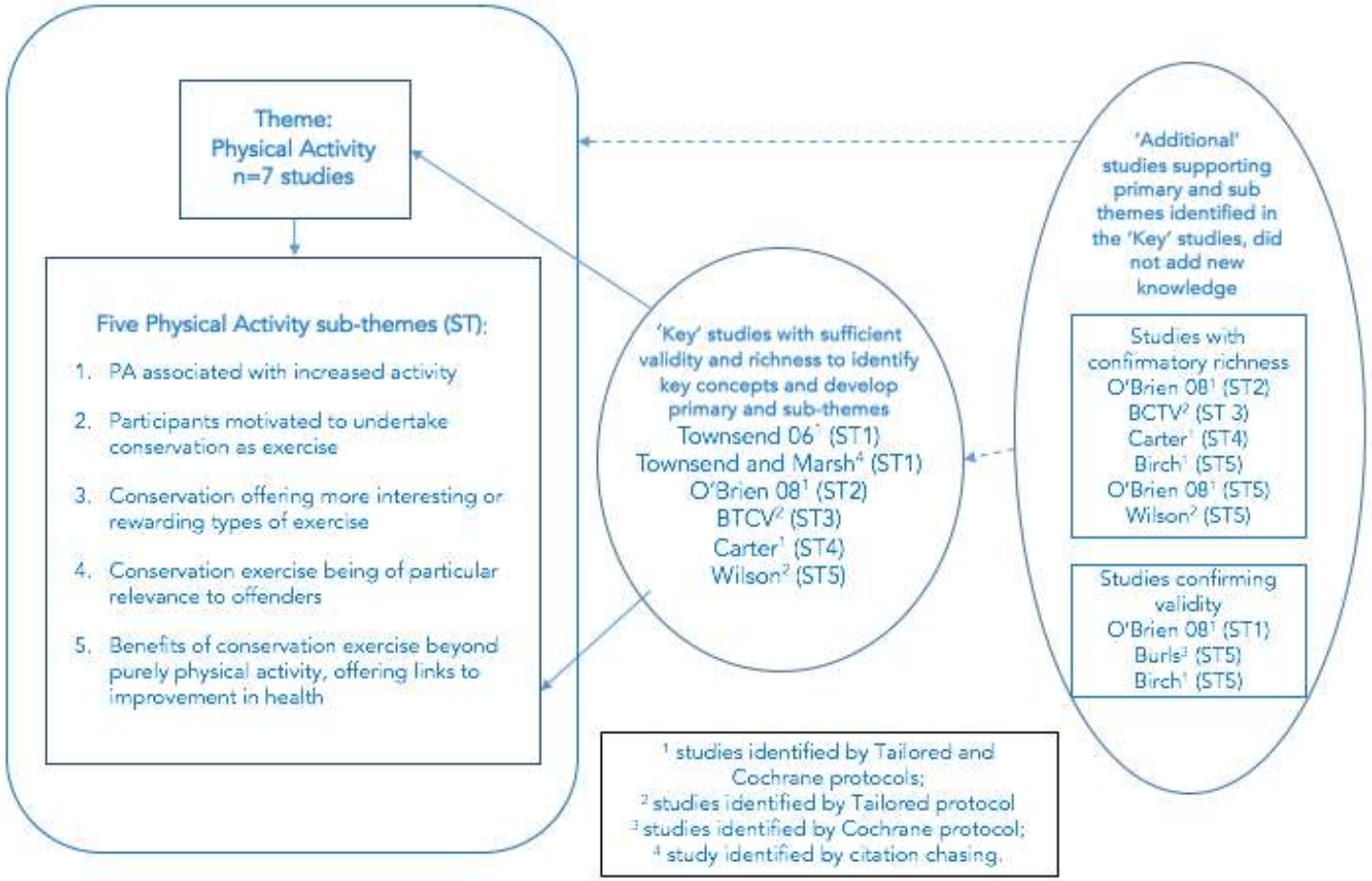

Figure 3: contribution of data to physical activity theme (qualitative studies)

- Assia (ProQuest);
- BIOSIS (ISI);
- British Education Index (ProQuest);
- British Nursing Index (ProQuest);
- CAB Abstracts (CAB Direct);
- Campbell Collaboration;
- Cochrane Public Health Specialized Register;
- DOPHER (EPPI);
- EMBASE (Ovid);
- ERIC (ProQuest);
- Global Health (Ovid);
- GreenFILE (EBSCO);
- HMIC (Ovid);
- MEDLINE in Process (Ovid);
- MEDLINE (Ovid);
- OpenGrey;
- PsycINFO (Ovid);
- Social Policy and Practice (Ovid);
- SPORTDiscus (EBSCO);
- TRoPHI (EPPI);
- Social Services Abstracts (ProQuest);
- Sociological Abstracts (ProQuest);
- The Cochrane Library (all via Wiley Interface);
- TRIP Database; and
- Web of Science (including conference citations index) (ISI).

Figure 4: databases searched