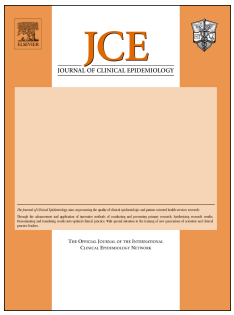
2017-04-28

Adjusting for unmeasured confounding in nonrandomized longitudinal studies: a methodological review

Streeter, Adam

10.1016/j.jclinepi.2017.04.022
Journal of Clinical Epidemiology
Elsevier BV

Accepted Manuscript



Adjusting for unmeasured confounding in non-randomised longitudinal studies: a methodological review

Adam J. Streeter, Nan Xuan Lin, Louise Crathorne, Marcela Haasova, Christopher Hyde, David Melzer, William E. Henley

PII: S0895-4356(16)30334-1

DOI: 10.1016/j.jclinepi.2017.04.022

Reference: JCE 9387

To appear in: *Journal of Clinical Epidemiology*

Received Date: 9 August 2016

Revised Date: 3 April 2017

Accepted Date: 24 April 2017

Please cite this article as: Streeter AJ, Lin NX, Crathorne L, Haasova M, Hyde C, Melzer D, Henley WE, Adjusting for unmeasured confounding in non-randomised longitudinal studies: a methodological review, *Journal of Clinical Epidemiology* (2017), doi: 10.1016/j.jclinepi.2017.04.022.

Adjusting for unmeasured confounding in non-randomised longitudinal studies: a methodological review

Adam J. Streeter[a,b]
Nan Xuan Lin[a,c]
Louise Crathorne[d]
Marcela Haasova[e]
Christopher Hyde[f]
David Melzer[g]
William E. Henley[a,*]

[a] Health Statistics Group, University of Exeter Medical School, University of Exeter, South Cloisters, St. Luke's campus, Exeter EX1 2LU, UK
[b] Biostatistics, bioinformatics & biomarkers group, Plymouth University Peninsula School of Medicine & Dentistry, University of Plymouth, Plymouth Science Park, Derriford, Plymouth PL6 8BX, UK
[c] Mathematics, physics & electrical engineering, Northumbria University, Sutherland Building, Newcastle-upon-Tyne NE1 8ST, UK
[d] Health Economics, University of Exeter Medical School, University of Exeter, South Cloisters, St. Luke's campus, Exeter EX1 2LU, UK
[e] Evidence Synthesis & Modelling for Health Improvement, University of Exeter Medical School, University of Exeter, South Cloisters, St. Luke's campus, Exeter EX1 2LU, UK
[f] Peninsula Technology Assessment Group, University of Exeter Medical School, University of Exeter, South Cloisters, St. Luke's campus, Exeter EX1 2LU, UK
[g] Epidemiology & Public Health, University of Exeter Medical School, RILD Building, RD&E Hospital Wonford, Barrack Road, Exeter EX2 5DW, UK

[*] Corresponding author: William Henley, University of Exeter Medical School, University of Exeter, South Cloisters, St. Luke's campus, Exeter EX1 2LU, UK
tel: +44 1392 726044
Email: w.e.henley@exeter.ac.uk

Abstract

Objective

Motivated by recent calls to use electronic health records for research, we reviewed the application and development of methods for addressing the bias from unmeasured confounding in longitudinal data.

Design

Methodological review of existing literature

Setting

We searched MEDLINE and EMBASE for articles addressing the threat to causal inference from unmeasured confounding in nonrandomised longitudinal health data through quasi-experimental analysis.

Results

Among the 121 studies included for review, 84 used instrumental variable analysis (IVA), of which 36 used lagged or historical instruments. Difference-in-differences (DiD) and fixed effects (FE) models were found in 29 studies. Five of these combined IVA with DiD or FE to try to mitigate for time-dependent confounding. Other less frequently used methods included prior event rate ratio adjustment, regression discontinuity nested within pre-post studies, propensity score calibration, perturbation analysis and negative control outcomes.

Conclusions

Well-established econometric methods such as DiD and IVA are commonly used to address unmeasured confounding in non-randomised, longitudinal studies, but researchers often fail to take full advantage of available longitudinal information. A range of promising new methods have been

developed, but further studies are needed to understand their relative performance in different contexts before they can be recommended for widespread use.

Keywords: method review, unmeasured confounding, unobserved confounding, longitudinal, observational data, electronic health records

Running title: Review of methods adjusting for unmeasured confounding in longitudinal data

Word count: 199

What is new?

What is already known

- Unmeasured confounding is a threat to the validity of observational studies based on data from non-randomised longitudinal studies

Key findings

- Longitudinal information that can be used to mitigate for unmeasured confounding in observational data is not always fully or properly utilised in health research.
- Instrumental variable analysis and difference-in-differences were the most commonly encountered methods to adjust for unmeasured confounding in a review of the health literature.
- There are a range of promising new methods, some of which utilise longitudinal information to relax the assumption of time-invariance for unmeasured confounders, but these are yet to be widely adopted.

What is the implication?

- All available methods rely on strong assumptions and more research is needed to establish the relative performance of different methods for particular problems and empirical settings.

1 Introduction

In the era of "big data" in medicine, the increasing availability of large, longitudinal patient databases is creating new opportunities for health researchers. A particular focus is on electronic health records (EHR) with routinely collected data collated from multiple care sites, often linked to external databases (e.g. death certificates). Built up over time, EHRs provide a sequential history of each patient's encounter with the healthcare system. Examples of EHRs include The Clinical Practice Research Datalink (CPRD), The Health Improvement Network (THIN), QResearch and ResearchOne in the UK, and the Kaiser Permanente Northern California Oracle Research Database in the US. The value of large medical data recorded for administrative purposes in national registries is already recognised [1,2], with the provision of funds to expand the adoption of EHRs in research for patient benefit in the US with the Health Information Technology for Economic and Clinical Health (HITECH) Act of 2009, and in the UK, with a consortium of funding bodies led by the Medical Research Council. Another important source of information for health care analysis is databases of insurance claims, such as Medicare in the US, and in this review we do not differentiate between EHRs and claims data.

A strength of EHRs and claims data is that they make it possible to study the comparative effectiveness of interventions and the associated risk of side-effects in a real-world setting. Although randomised trials provide the gold standard of evidence, observational studies based on observational patient databases offer the potential to study more patients from a wider variety of risk groups with a longer follow-up period at a fraction of the cost. However, in the absence of randomisation, selection for treatment is often knowingly based on specific characteristics, such as frailty, disease severity or the risk of an outcome. If the indication for treatment is also related to prognosis, confounding by indication arises leading to biased estimation of effectiveness. There is a large pharmacoepidemiologic literature on this topic and current best practice is to use design-based approaches such as the Active Comparator, New User Design to help mitigate bias where possible[3].However, residual differences between the treatment arms other than the treatment itself may still confound the intervention effect under study whether or not such an approach is used. If the confounding variables are both known to the study investigators and measurable, then these could potentially be adjusted for in prospective non-randomised studies. With retrospectively recruited subjects, however, the recording of such variables is outside the control of the investigator. Analyses of non-randomised studies that fail to account for relevant confounders may have important negative consequences for health policy and patient safety.

Methods described as the quasi-experimental (QE) approach[4], can be deployed to account for confounding by unobservable characteristics. These do not attempt to directly adjust for resulting bias, but use available information to achieve this indirectly under certain conditions and assumptions. The aim of this systematic review is to review current practices in dealing with unmeasured confounding in individual-level longitudinal health data and to capture methodological developments in this area. While previous systematic reviews have been conducted to look at use of propensity score methods for measured confounders [5,6], we are unaware of any systematic review comparing use of methods for addressing unmeasured confounding in non-randomised, longitudinal data. We were particularly interested in how an individual's history could be leveraged to evaluate the effects of unmeasured confounding and how the extra longitudinal information could be incorporated to improve adjustment for confounding bias. We intend for this review to contribute to the development of best practice in addressing unmeasured confounding in longitudinal data. The results should therefore help inform researchers intending to utilise "big data" from electronic health records.

2 Methods

2.1 Search strategy

Our search strategy was informed by, but not limited to, known methods for addressing unmeasured confounding. The search strategy is recorded in Appendix A. The following electronic databases were searched: MEDLINE (via OvidSp including In-Process & Other Non-Indexed Citations) and EMBASE (via OvidSp 1996 to 2015 Week 21). We included all citation dates from database inception to May 2015. All references were exported into Endnote X7 (Thomson Reuters).

2.2 Inclusion and exclusion criteria

The review included any non-randomised comparative studies that sought to adjust for unmeasured confounding in longitudinal data with repeated observations on identifiable individuals. In the interests of good practice, eligible papers had to explicitly identify the problem of bias arising from the selection on unobservable characteristics in the data, rather than routinely apply a QE design without this justification. For estimates of comparative effectiveness, eligible studies had to have independent control arms for each treatment of interest. Therefore, single arm studies were

excluded. Studies based on case-only designs, including the case-crossover design and the self-controlled case-series design, in which confounding is controlled by making comparisons between exposed and unexposed periods for the same individual were also excluded. Observational studies were not excluded based on the exposure under study so studies into the effects of passive exposures (medical conditions, environmental exposures etc) were included alongside studies of both the intended and adverse effects of active interventions. We note that good proxies for unmeasured confounding, or observed variables that sufficiently describe a latent variable such as frailty, would be preferable to dealing with the bias resulting from unmeasured confounders. If suitable proxies are identified and recorded, then there are in effect no unobserved confounders and the proxies could simply be adjusted for in the analysis, obviating the need for methods to adjust for the unobserved confounders. For this reason, adjustments for proxies of unmeasured confounders, including high-dimensional propensity scores, did not fall within the scope of this study. To be consistent with the "big data" theme of EHRs, a minimum sample size of 1000 participants was applied. This also set a minimum condition for the application of Instrumental Variable (IV) and Regression Discontinuity (RD) designs stipulated in the Quality of Effectiveness Estimates from Non-randomised Studies (QuEENS) checklist. Finally, we only accepted analyses of individual level data. We were aware that some studies may use analytical methods, such as difference-in-differences that aggregate the data at a treatment-group level. We therefore only included those studies, in which the same patients could be tracked over the time-frame of the sample. Conversely, some methods, such as instrumental variable analysis, make no explicit demands for longitudinal data at the patient level. However, we included such studies where the sample was based on the availability of patient-level longitudinal information, with a history possibly but not necessarily preceding the time of exposure. We did not discriminate between data sources, as patient-level data will often arise from medical insurance claims in the US, as opposed to clinically-purposed databases in other countries.

Only studies written in English were included.

The following publication types were excluded from the review:

- systematic reviews of primary studies.
- randomised controlled trials
- cross-sectional data
- preclinical and biological studies
- narrative reviews, editorials, opinions

2.3 Study selection

Studies retrieved from the searches were selected for inclusion through a two-stage process according to the inclusion/exclusion criteria specified above. First, abstracts and titles returned by the search strategy were screened for inclusion independently by two researchers. In case of doubt, the article in question was obtained and a subsequent judgement on relevance was based on the full article. Disagreements were resolved by discussion, with involvement of a third reviewer when necessary. Following the initial screening, full texts of identified studies were obtained and screened firstly by a single reviewer. In case of doubt, a second reviewer decided on the suitability of a paper. Where multiple publications of the same study were identified, data were extracted and reported as a single study.

2.4 Evidence synthesis

The details of each study's design and methodology and the key characteristics of the data source were tabulated and discussed. We present a summary of the methods we found that can mitigate for confounding, or its synonyms as unmeasured, unobserved, hidden or residual. We note the historical frequency and context of the application of those methods, to comment on progress in causal inference and identify directions for future research.

3 Results

3.1 Included studies

Our searches returned 734 unique titles and abstracts, with 275 papers retrieved for detailed consideration Of the 275 studies eligible for a full-text review, 154 were excluded (see flow diagram: Figure 1).

A total of 121 studies were identified as performing a QE analysis on non-randomised longitudinal data on human subjects, identifiable at an individual level, and so included for a full review of the text (Appendix B).

The QE methods identified in the review are summarised inTable 1. The most frequent method was instrumental variable analysis (IVA) found in 86 of the studies (Figure 2) – a method that uses an unconfounded proxy for the intervention or exposure. For successful adjustment, the proxy or instrument should be strongly, causally associated with the exposure or intervention, and the

instrument should only affect the outcome through the exposure. In addition to IVA, three of these also applied difference-in-differences (DiD) – a method that typically uses pre-exposure outcomes to adjust for unmeasured confounding and assumes any trends unrelated to the exposure are the same in both groups. Seven more studies derived estimates from a combination of both IVA and DiD, two of which assumed an absence of higher order autocorrelation to use lagged observations of the treatment variable as an instrument. Beside the 11 studies applying DiD either in conjunction with or in addition to IVA, we identified a further 21 studies, in which the sole QE method was recognised as a DiD approach.

We found five studies applied the prior event rate ratio method, a before-and-after approach that can be aggregated to the treatment level for survival or rate outcomes and analogous to DiD. In all five cases the methods were applied to longitudinal, individual patient data. Similarly regression discontinuity (RD) was used for such data in three of the studies included for review. Another three focused on propensity score calibration (PSC). One study introduced perturbation testing and perturbation analysis, while another discussed the use of negative control outcomes.

3.1.1 Studies excluded at full text

The principal reason for exclusion in 94 of the studies, according to our eligibility criteria, was the absence of longitudinally observed, non-randomised outcomes on all individually identifiable persons, although other characteristics may also have justified their exclusion. No particular method was associated with the absence of longitudinal data on identifiable individuals with this studies in this exclusion category comprising 59% DiD and 28% instrumental variable analyses compared, respectively, to 53% and 32% of all 154 of the rejected studies. Having fewer than 1000 longitudinally observed individuals excluded 23 studies, among which those using instrumental variable analysis (IVA) numbered 15. Seven were excluded for not employing a QE method for unmeasured confounding. Five studies presented exploratory analyses without a focused clinical question; five were either method reviews or commentaries without an application of methods to data; one study duplicated a dataset already marked for inclusion, while another failed to specify the instrumental variable used. Of particular note were the 18 studies using the DiD approach that were excluded because no explicit justification was made for using the method to address unmeasured confounding, or any of its synonyms. In these studies, justification of the method was centred more on econometric concerns over time trends, and presented in terms of controlling for those trends rather than pre-existing differences between the control and exposed group.

3.2 Results of the included studies

So far studies have been categorised according to their identified QE method. However, certain properties are shared across some of the methods, and can be classified according to how they reconcile their specific assumptions with the information offered by the structure of big, longitudinal data that typifies EHRs. In particular, we organised our results around how each method had incorporated longitudinal information, and the assumptions required. The stable of before-and-after methods, that includes PERR and DiD, implicitly incorporates longitudinal information. Thereafter the challenge is how to relax the assumption of time-invariant confounding. Conversely, IVA is not uniquely applicable to longitudinal data, but we were able to broadly classify the types of instruments used (Table 2), some of which did utilise longitudinal information. We found out of the total 121 studies, 77 incorporated some element of longitudinal information into their analysis.

3.2.1 Incorporation of external/additional data

The propensity scores (PS), the predicted probability of exposure or treatment conditioned on measured confounders,were used in the seminal work on propensity score calibration (PSC) by Stürmer to calibrate an error-prone PS against a gold-standard PS and hence arrive at an inference for the level of unmeasured confounding bias [7]. The two subsequent PSC papers examined the tenability of the method's assumptions, firstly using simulated data to evaluate the conditions necessary to violate the surrogacy assumption [8]. The second primarily used simulated data and applied the results to registry data to demonstrate a framework for determining size and direction of bias from one measured and one hidden confounder [9].

3.2.2 High-dimensional data

Since PSC collapses multiple, potential confounding variables down to the single dimension of a propensity score, the three PSC papers can also be considered a means of dealing with high-dimensional data. In addition to these, our review also included a novel data-mining approach that proposed to exploit the many factors (perturbations) that may be weakly associated with the unmeasured confounders from a high dimension dataset [10], for which longitudinal data may mitigate for incorrect adjustment of a collider. Perturbation analysis was successfully demonstrated on simulated data, although accidental inclusion of a measured confounder required many more

perturbations to correct the resulting bias. Both the perturbation method and PSC were also proposed as sensitivity analyses.

3.2.3 Quasi-experimental adjustment without longitudinal assumptions

Those studies characterised as using a QE method without any longitudinal dimension were PSC and PT as described above. We also added to this category 11 examples of Mendelian IVA [11–21] plus 32 other IVAs without historic or lagged instruments [22–53]. While time-based instruments may at first seem longitudinal, these instruments, such as date of therapy, would need to be related to previous exposures or outcomes to be considered longitudinal. In some cases, survival times or rate data were used, but such outcomes do not intrinsically imply longitudinal adjustment for confounding. In spite of these "cross-sectional" approaches, all studies were based on some form of longitudinal data at the person level, as demanded by our inclusion criteria. Among the 43 non-Mendelian IVA papers in this non-longitudinal category, one study adjusted for non-longitudinal fixed effects within twins [39]. In another three, discussed below, the analysis was supplemented with DiD [38,47], and with IVA applied to first-differences[54].

One study examined the effect of lagged, cumulative exposure to radiation on lung cancer in uranium miners and nuclear workers [55]. The problem of unmeasured confounding was addressed using a method developed in earlier work that proposed negative control outcomes and exposures as a means of both detecting and potentially resolving confounding bias[56]. Here the choice of death due to chronic obstructive pulmonary disorder as a negative control outcome was informed by clinical knowledge of there being no direct relationship with the exposure except through the possible confounder, smoking. Given a plausible negative control outcome or exposure, the method offers at least a means of testing for confounding, and potentially a method of adjustment under the assumption that the association between the unmeasured confounder and the negative outcome is similar in magnitude to that between the same confounder and the outcome of interest.

3.2.4 Quasi-experimental adjustment assuming time-invariant longitudinal information

We found 36 IVA studies that used lagged information or history about the individuals' exposure as instruments [54,57–92]. One study had recourse to the random assignment from a previous study, and used this as an instrument [69]. Except for that and four other different exceptions, the instruments were all based at least in part on the previous intervention, or history of interventions, of the clinician or healthcare facility. Characteristics of the clinician or facility may be chosen as

instruments as they are more likely to affect the treatment only. This avoids direct associations with the individual and their outcome, and so better enforces the exclusion restriction – the exclusion of the instrument's association with the outcome except through the treatment under study. While no assumptions are made about the dependence of confounding on time, the strength of the instrument clearly rests on a significant association between previous treatment(s) and the current treatment under investigation. In this regard, if the strength of an instrument varies with time, this may undermine its utility.

In total, 24 studies also incorporated longitudinal information through the stable of methods that, in an abuse of terminology, we collectively referred to as the DiD approach. These included the 18 examples cited as using DiD regression [93–110] alone, and four fixed effects (FE) [111–114]. Either through fixed effects at the individual level or through aggregate-level regression operationalizing the DiD approach, these methods "ignore" the effect of confounding, which is assumed to be time-invariant. At the individual level, time invariant confounding can be ignored by assigning nuisance dummy variables for each individual, or cancelled out through demeaning the observations, or through the first differences of observations on each individual. Two of the studies also extended DiD to allow different exposure effects and trends across two-level sub-groups in the higher-order contrast of difference-in-difference-in-differences [95,106]. Fourteen studies also adjusted for individual-level fixed effects either through direct inclusion of their covariates, or through matching or weighting on the propensity score of the covariates. This was perhaps a more rigorous and precise approach, accounting for known confounders, and yielding smaller standard errors for the estimated treatment effect. However, an assumption of time-invariant confounding was still required, with a null difference between exposure groups in the prior period being evidence of adjustment for time-invariant confounding only. Two of the 24 DiD studies also re-analysed their data using IVA [38,47], which provided an albeit limited opportunity to compare the relative performance of these methods. In the study by Schmittdiel et al. of how statins delivered by mail order affects cholesterol control[47], the intervention coefficient from modelling the single main outcome was larger through DiD analysis and its standard error smaller than those from IVA, large standard errors being a feature of weak instruments. The study by Lei and Lin investigated the effect of exposure to a new medical scheme on 15 health outcomes and rates of health-service utilisation[38]. The effects were either not significantly different from the null or were significant and of similar magnitude with similar standard error except for two outcomes, where the effect size was significantly larger for IVA.

Time-invariant confounding, also known as the parallel trends assumption, was relaxed by including dummy variables for the year and its interaction with the treatment dummy in a fixed-effects analysis, which allowed the unobserved trend to vary between exposure groups [113] using methods developed in economics and therefore not captured by this review [115,116]. The results from this DiD with differential trend model were presented alongside those from the simple pooled DiD model and DiD with individual fixed-effects for the effect of financial incentives in care services. Tests confirmed parallel trends could be assumed in three outcomes, but out of the five outcomes presented, four were statistically significant and in all, the estimated effect size by differential trends was greater.

Our review also included six studies applying the prior event rate ratio method, a before-and-after analogue applicable to survival and rate data [117–122]. The first two published were the seminal presentation of the method applied to registry data. Also included was a comprehensive evaluation by Uddin et al. of the performance of PERR under a wide array of simulated, theoretical settings, under which bias was shown to increase with a greater effect of the prior events on subsequent exposure or intervention. When prior events strongly influence the likelihood of treatment, the exposure effect from the PERR method can be more biased than estimates from conventional methods[121]. The problem was re-examined in a recently published study, which provided a more general statistical framework for PERR adjustment and considered the potential for generalising the method to allow more flexible modelling[122].

3.2.5 Dynamic, longitudinal quasi-experimental methods and time-varying information

While regression discontinuity (RD) could suggest a longitudinal design, this is not exclusively so, and two RD studies were excluded because of this (one applied to spatial data while the other data was not longitudinal). Of those included all three could be said to accommodate time varying trends [123–125], and two of these were nested within a pre-post design: Zuckerman et al. were explicit in their methodological study in identifying the robustness to time-varying confounding, in which inhaler use in asthmatic patients was served as both the outcome variable in the post-test period as well as the assignment variable in the pre-test period[125]. In the study of the effect school-leaving age on mortality by Albouy, different slopes were modelled for the assignment variable, year of birth, after the cut-off date[123]. This acknowledged different maturation rates after assignment. However, as long as the assumptions of the method were met, assignment should have been as good as randomised, and so no further assumptions about the temporality of confounding was required.

We also picked up six examples where IVA had been combined with either DiD or a fixed effects model, first appearing in our review with example from 2003 [126]. In Fortney's 2005 study of treatment for depression [127], this combination method was justified as a control for time varying confounding, referred to as second-order endogeneity. Further examples of the fixed-effects instrumental variable model were found [128,129]. The roles of lagged treatments and outcomes as possible IVs and predictors were extensively considered in O'Malley's study of whether the introduction of more expensive medication could have led to improved cost-effectiveness in the long term[54]. The author cautioned that the exclusion restriction may be difficult to satisfy when using the lagged treatment as an IV after first differencing. However, two studies [130,131] used differences in the lagged explanatory variable as the IVs to adjust for second-order endogeneity in a first-differences analysis following methods, not captured by our review, but developed in the realm of Economics [132–134]. Referred to as the dynamic panel model or IV-GMM, this method was implemented efficiently through generalised method of moments. In their report on healthcare expenditure in patients with rheumatoid arthritis, Kawatkar et al. found the yielded estimates were further from the null with larger standard errors when compared to those from FE alone[130].

3.3 Implementation of methods

While choice of method in each study often rested on which extra information was available to address the issue of unmeasured confounding, method selection may also have been informed by the research area. The negative control method had its origins in epidemiology, with applications to occupational health policy. Likewise, the PERR method was developed exclusively on health data, with applications to drug safety and public health policy. Reflecting their origins in health econometrics, some studies were published in journals partially or entirely dedicated to the subject, with 15 published[38,54,93–95,98,103,104,111–114,126,127,130] in this field out of the 32 studies using DiD and 29[23,24,28–30,32,33,36,41,46,48,49,51,52,66,69–72,77,81,84,86,135] out of the 86 using IVA. Under the inclusion criteria, all studies had health outcomes or interventions. Mendelian IVA necessarily includes genetic information, and all were published in health-related journals. In contrast, all three studies using RD were published in health econometric journals.

Before implementing one of the proposed methods, a natural first step is for the researcher to try to assess how much bias from unmeasured confounding is likely to be present. While many of the included studies reported raw or unadjusted descriptive estimates, bias estimation was limited

either to considering the contribution from known confounders, including those summarised as a propensity score, or to methods, such as perturbation testing/analysis and negative controls methods, in which bias evaluation is an incremental step in adjustment. Under the assumption of time-invariant confounding, the difference-in-differences method may potentially offer a way of evaluating bias by modelling group differences in the pre-exposure period. However, few studies evaluated hidden bias in this way[47,96,112]. The regression formulation of the DiD method effectively by-passes separate analysis of the prior period. Instead studies often discussed the within-group changes over time. Similarly, the prior-period estimate from the PERR method implicitly offers an evaluation of confounding bias under the same assumptions, yet none of the studies presented information on outcomes in the prior period in this way. A direct evaluation of unmeasured confounding is less straight-forward in IVA, with further diagnostic tests only recently developed for the association between instrument and confounders[136,137] .

4 Discussion

This review examined the application of methods to detect and adjust for unmeasured confounding in observational studies, and was motivated by recent calls to utilise EHRs. Most of the reviewed studies used more established methods such as DiD and particularly IVA. We summarised how studies exploit the longitudinal information afforded by EHRs.

It may be tempting to view electronic health records and medical insurance claims data as a problem of large observational data, and hence search for solutions through data mining. However, ethics governing patient data collection, plus limited clinician time is likely to preclude data with very large dimensions. For that reason, it is doubtful there would be enough dimensions for a method like Perturbation Analysis (PA) to be a practical solution. In addition, a greater number of variables would likely include enough information about the confounders to obviate the need for further adjustment through PA. More generally, the purpose of EHRs primarily as an administrative tool limits the scope for data mining of known confounders. Similarly, limited availability of gold-standard datasets may have confined the use of external data, as in PSC, to but a few examples.

We were surprised by the number of studies using IVA alone. While Mendelian randomisation has its advantages for many studies as a reasonable guarantor of the exclusion restriction, in general IVA typically suffers from the weak-instrument problem, resulting in large standard errors and wide confidence intervals. Longitudinal data offer an opportunity to reinforce the exclusion criteria

by choosing historical or lagged instruments. However in doing so, the causal structure needs to be understood to avoid opening up "back door" paths and inducing further bias[54]. DiD arguably offers advantages over IVA in being more intuitive and easier to conceptualise, and with the longitudinal data in EHRs it should be inherently easier to work with prior observations than to identify strong instruments. Even though before-and-after methods are not subject to the imprecision of weak instruments, the resulting estimates are only unbiased if the unobserved confounders exert a constant effect over the observation windows before and after exposure. Where multiple observations per individual exist, time may be paramaterised and different trends between exposure groups can be accommodated in DiD with differential trends, but a time invariant assumption about confounding must still be made. To partially or wholly relax this particular assumption, instrument variable analysis can be incorporated into the fixed effects model. Assuming the instrument's exclusion restriction is satisfied then this doubly-robust approach affords the advantage of DiD over possibly weak instruments, while mitigating for some or all of the time-dependent confounders ignored by DiD alone. Similarly, where multiple previous treatments or exposures are recorded, the differenced lagged treatments can be utilised as IVs in a fixed effects model to accommodate time-dependent confounding bias using the generalized method of moments system, referred to as IV-GMM or the dynamic panel model.

Another potentially robust approach to unmeasured confounding would the RD design, although the small number of examples in our review probably reflects the limited number of scenarios where this can be reasonably applied. Another concern over and above the usual technical challenges of applying the RD method is that in spite of heath records promising ample data, the sample would need to be reduced to an interval around the cut-off that ensures exchangeability of the two treatment groups. In this case generalisability would be restricted to individuals with characteristics found in the interval. As with RD, PERR was another method that was found in relatively few studies. This may have been in large part due to its recent development, rather than any technically demanding aspect of its application, since it simply extends the before-and-after approach of DiD to survival and rate data - outcomes that are common enough in health research. However, the PERR approach does require strong assumptions including time-invariant confounding and the absence of an effect of prior events on likelihood of future treatment[122].

Methods such as IVA and DiD have their origins in the sphere of econometrics, where randomised experiments are rare. We found that in importing DiD, some of the studies failed to explicitly acknowledge the problem of confounding bias. Instead justification for the method was presented in terms of the common trends assumption. Discussion of possible confounding bias is regarded as

essential by most QA toolkits for observational data, and it is important that health researchers explicitly recognise this threat to the internal validity of non-randomised studies. Conceptually a non-temporal analogue of DiD would be the NCO method, which itself was presented foremost as a method for detecting unmeasured confounding. Given doubts over satisfying necessary assumptions for their implementation, authors of this method along with propensity score calibration and perturbation analysis have suggested that, as sensitivity analyses, these can at least offer an insightful complement to QE adjustment.

Choosing between methods to reduce unmeasured confounding bias is challenging and we found few studies that directly compare methods. The performance of different methods will depend on factors such as the nature of the underlying confounding, the type of exposure and outcome, and the sample size[138] The type of data available will also guide the choice of method. For example, the instrumental variable method requires a suitable instrument and DiD / PERR require data on at least two periods. In practice, no one method is likely to be best suited to all problems, and it is essential for investigators to carefully assess the potential biases in each proposed study, where possible tailoring the methods or combination of methods to address these biases[139]. Our review has highlighted how use of longitudinal information is one additional and potentially important consideration in this process.

While our review focussed on the problem of adjustment using analytic methods, many problems associated with observational data may be pre-empted by use of an appropriate study design[140]. Before choosing an appropriate analytic method, it is recommended that investigators carefully identify and match individuals for the control and intervention groups in order not to exacerbate any bias[3]. The importance of study design is often discussed with a view to minimising confounding bias from unmeasured sources, with the subsequent adjustment accounting for observed confounders only[141], usually through the matching, weighting or adjustment of propensity scores[142]. Where the success of the design remains in doubt, or its criteria cannot be fully met, then investigators will inevitably need recourse to some of the alternative methods reviewed in this report.

The reviewed studies did not seek to distinguish between the different mechanisms of bias. Confounding by indication, deemed intractable by many researchers using the observed data[143], was seen to create additional sources of bias in two separate simulation studies applying the "longitudinal" method of PERR, when an association was modelled between prior events and treatment status in the study period[121,122]. Another common form of selection bias in

pharmacoepidemiologic studies is the healthy user bias and this works in the opposite direction to confounding by indication, distorting treatment-outcome associations towards the treatment looking beneficial[3]. Further research is needed to understand how each of the methods in this review is affected by the different types of confounding.

An inherent limitation of this large, wide-ranging review is that it precluded meaningful data synthesis due to the mix of different data and study types. Furthermore, we could only find a few examples where the performance of different methods was compared within the same study. We also stipulated in the inclusion criteria that unmeasured confounding, or any of its synonyms, should be given as justification for methods in its adjustment. This may have inadvertently excluded some papers, where justification was implicit, but good practice in health research demands acknowledgement of this source of bias where applicable. While our search terms were specific to the scope of our review, we accept that this may have inadvertently excluded relevant methods and studies. Some methods, such as negative control outcomes, that were identified in the original search were not included as explicit terms in the search strategy, and further secondary searches may have uncovered additional studies using these methods. We also acknowledge that there may be other relevant methods for addressing unmeasured confounding that have been missed by the search strategy. Consequently, we made inferences about the relative application of methods with caution. However, we were surprised so many studies focussed solely on IVA as the sole means of adjustment. A similar conclusion was echoed by a different review on regression discontinuity designs that found interest was growing in RD only as recently as 2014 [144].

By choosing to focus on methods with an independent control arm for each treatment, our review excluded case only designs including case-crossover designs (CCO) and the self-controlled case-series design. This class of methods addresses unmeasured confounding by making comparisons within individuals so that each individual acts as his or her own control. Another case-only design, the case-time control design, is an extension of the CCO design that uses information from a historical control group in a similar way to the PERR method. These approaches are reviewed by Uddin et al[138] and Nordmann et al[145].

This review has considered a range of promising new methods for addressing unmeasured confounding in non-randomised studies. However, consistent with prior research on dissemination and uptake of statistical innovations[146], the rate of knowledge translation has been slow and we found that most studies in our review used established methods such as IVA and DiD. A recent study by Cadarette et al has shown how Rogers' Diffusion of Innovations model can be used to

describe the adoption of novel methodologies in pharmacoepidemiology[147] and this provides a useful resource for interpreting the uptake of methods in this review. Cadarette et al proposed five principles for authors of methodological innovations that may improve translation into practice [147]: (1) clearly describing the methods using foundational principles; (2) comparing results to established methods; (3) providing sample data, code or calculation examples; (4) early communication, support and testing; and (5) providing methodological and reporting guidance. These recommendations offer a useful checklist for researchers developing methods for addressing unmeasured confounding in observational studies. Of particular relevance in the context of this review is the need for more extensive evaluation and comparison of the emerging methods in a range of settings. The review also addresses the need for methodological guidance through highlighting the potentially important role of longitudinal information in addressing confounding bias and has identified this as an area for further development.

5 Conclusions

Our review showed how seminal work in econometrics has influenced practice in dealing with unmeasured confounding in clinical and epidemiological research. Although the issue of unmeasured confounding is widely acknowledged, we found that longitudinal information in observational studies appears under-utilised. Lagged and historical characteristics associated with the treatment may help enforce the exclusion restrictions of instrumental variables under the appropriate causal structures, while before-and-after methods, such as DiD and PERR, afford an intuitive approach without the imprecision of weak instruments. Furthermore, they offer a direct evaluation of time-invariant confounding bias. The most robust methods we found applied instrumental variable analysis to the fixed effects difference-in-differences method, where such suitable instruments or difference lagged variables could be assumed to satisfy the exclusion restriction. While there are sometimes good technical reasons for choosing one mode of analysis over another, many questions remain over the most appropriate methods. All methods rely on assumptions, but little guidance is available to applied researchers as to the empirical settings in which particular methods can be safely used. Few studies directly compare different methods and more research is needed to the establish the relative performance of the methods in realistic settings.

Acknowledgements

This research was funded by the Medical Research Council [grant number G0902158] with additional support from the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care (CLAHRC) for the South West Peninsula and

592 the National Institute for Health Research (NIHR) School for Public Health Research Ageing
593 Well programme. The School for Public Health Research (SPHR) is funded by the National
594 Institute for Health Research (NIHR). SPHR is a partnership between the Universities of
595 Sheffield, Bristol, Cambridge, UCL; The London School for Hygiene and Tropical Medicine; the
596 University of Exeter Medical School; the LiLaC collaboration between the Universities of
597 Liverpool and Lancaster and Fuse; The Centre for Translational Research in Public Health, a
598 collaboration between Newcastle, Durham, Northumbria, Sunderland and Teesside
599 Universities. The funders had no role in study design, data collection and analysis, decision to
600 publish, or preparation of the manuscript. The views expressed in this publication are those
601 of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

References

1. Murdoch TB, Detsky AS. The inevitable application of big data to health care. *JAMA*. 2013;309(13):1351-1352. doi:10.1001/jama.2013.393.
2. Safran C, Bloomrosen M, Hammond WE, et al. Toward a national framework for the secondary use of health data: an American Medical Informatics Association White Paper. *J Am Med Inform Assoc*. 2007;14(1):1-9. doi:10.1197/jamia.M2273.
3. Lund JL, Richardson DB, Stürmer T. The Active Comparator, New User Study Design in Pharmacoepidemiology: Historical Foundations and Contemporary Application. *Curr Epidemiol Reports*. 2015;2(4):221-228. doi:10.1007/s40471-015-0053-5.
4. Cook TD, Campbell DT. *Quasi-Experimentation: Design & Analysis Issues for Field Settings*. 3rd ed. Chicago: Rand McNally, 1979; 1979. http://lib.exeter.ac.uk/record=b1090074~S6. Accessed January 13, 2015.
5. Shah BR, Laupacis A, Hux JE, Austin PC. Propensity score methods gave similar results to traditional regression modeling in observational studies: a systematic review. *J Clin Epidemiol*. 2005;58(6):550-559. doi:10.1016/j.jclinepi.2004.10.016.
6. Stürmer T, Joshi M, Glynn RJ, Avorn J, Rothman KJ, Schneeweiss S. A review of the application of propensity score methods yielded increasing use, advantages in specific settings, but not substantially different estimates compared with conventional multivariable methods. *J Clin Epidemiol*. 2006;59(5):437-447. doi:10.1016/j.jclinepi.2005.07.004.

56. Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative Controls. *Epidemiology*. 2010;21(3):383-388. doi:10.1097/EDE.0b013e3181d61eeb.

115. Wagstaff A, Moreno-Serra R. Europe and central Asia's great post-communist social health insurance experiment: Aggregate impacts on health sector outcomes. *J Health Econ*. 2009;28(2):322-340. doi:10.1016/j.jhealeco.2008.10.011.

116. Bell B, Blundell R, Reenen J Van. Getting the Unemployed Back to Work: The Role of Targeted Wage Subsidies. *Int Tax Public Financ*. 1999;6(3):339-360. doi:10.1023/A:1008787013977.

132. Arellano M, Bond S. Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *Rev Econ Stud*. 1991;58(2):277. doi:10.2307/2297968.

133. Arellano M, Bover O. Another look at the instrumental variable estimation of error-components models. *J Econom*. 1995;68(1):29-51. doi:10.1016/0304-4076(94)01642-D.

134. Blundell R, Bond S. Initial conditions and moment restrictions in dynamic panel data models. *J Econom*. 1998;87(1):115-143. doi:10.1016/S0304-4076(98)00009-8.

135. MacKenzie TA, Tosteson TD, Morden NE, Stukel TA, O'Malley AJ. Using instrumental variables to estimate a Cox's proportional hazards regression subject to additive confounding. *Heal Serv Outcomes Res Methodol*. 2014;14(1-2):54-68. doi:10.1007/s10742-014-0117-x.

136. Jackson JW, Swanson SA. Toward a clearer portrayal of confounding bias in instrumental variable applications. *Epidemiology*. 2015;26(4):498-504. doi:10.1097/EDE.0000000000000287.

137. Davies NM. An even clearer portrait of bias in observational studies? *Epidemiology*. 2015;26(4):505-508. doi:10.1097/EDE.0000000000000302.

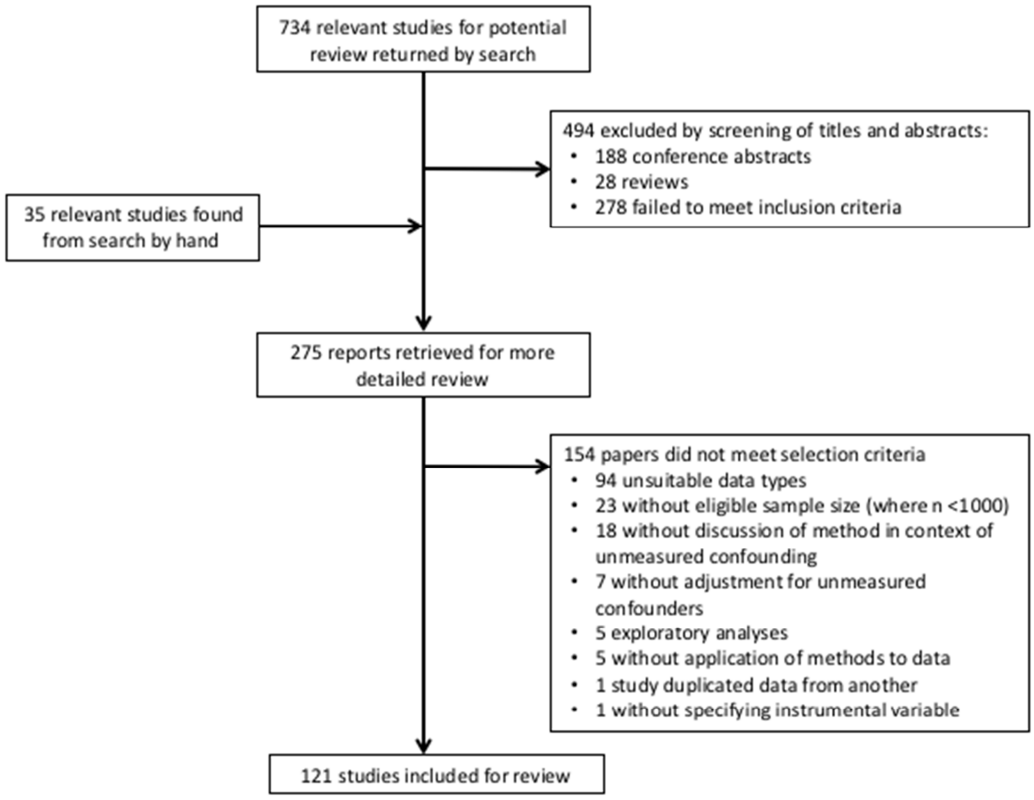
138. Uddin MJ, Groenwold RHH, Ali MS, et al. Methods to control for unmeasured confounding in pharmacoepidemiology: an overview. *Int J Clin Pharm*. April 2016. doi:10.1007/s11096-016-0299-0.

139. Alemayehu D, Alvir JMJ, Jones B, Willke RJ. Statistical issues with the analysis of nonrandomized studies in comparative effectiveness research. *J Manag Care Pharm*.

2011;17(9 Suppl A):S22-6. http://www.ncbi.nlm.nih.gov/pubmed/22074671. Accessed July 18, 2014.
140. Velentgas P, Dreyer NA, Nourjah P, Smith SR, Torchia MM. *Developing a Protocol for Observational Comparative Effectiveness Research: A User's Guide*. Agency for Healthcare Research and Quality (US); 2013. http://www.ncbi.nlm.nih.gov/pubmed/23469377. Accessed October 25, 2016.
141. Freemantle N, Marston L, Walters K, Wood J, Reynolds MR, Petersen I. Making inferences on treatment effects from real world data: propensity scores, confounding by indication, and other perils for the unwary in observational research. *Bmj*. 2013;347(nov11 3):f6409-f6409. doi:10.1136/bmj.f6409.
142. Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behav Res*. 2011;46(3):399-424. doi:10.1080/00273171.2011.568786.
143. Bosco JL, Silliman RA, Thwin SS, et al. A most stubborn bias: no adjustment method fully resolves confounding by indication in observational studies. *J Clin Epidemiol*. 2010;63(1):64-74. doi:http://dx.doi.org/10.1016/j.jclinepi.2009.03.001.
144. Moscoe E, Bor J, Bärnighausen T. Regression discontinuity designs are underutilized in medicine, epidemiology, and public health: a review of current and best practice. *J Clin Epidemiol*. 2015;68(2):122-133. doi:10.1016/j.jclinepi.2014.06.021.
145. Nordmann S, Biard L, Ravaud P, et al. Case-Only Designs in Pharmacoepidemiology: A Systematic Review. Little J, ed. *PLoS One*. 2012;7(11):e49444. doi:10.1371/journal.pone.0049444.
146. Pullenayegum EM, Platt RW, Barwick M, Feldman BM, Offringa M, Thabane L. Knowledge translation in biostatistics: a survey of current practices, preferences, and barriers to the dissemination and uptake of new statistical methods. *Stat Med*. 2016;35:805-18.
147. Cadarette SM, Ban JK, Consiglio GP, Black CD, Dubins D, Marin A, Tadrous M. Diffusion of Innovations model helps interpret the comparative uptake of two methodological innovations: co-authorship network analysis and recommendations for the integration of novel methods in practice. *J Clin Epidemiol*. 2016 Dec 23. pii: S0895-4356(16)30835-6.

686 **Figures**

687



688

689 **Figure 1:** Flow diagram for method review

690

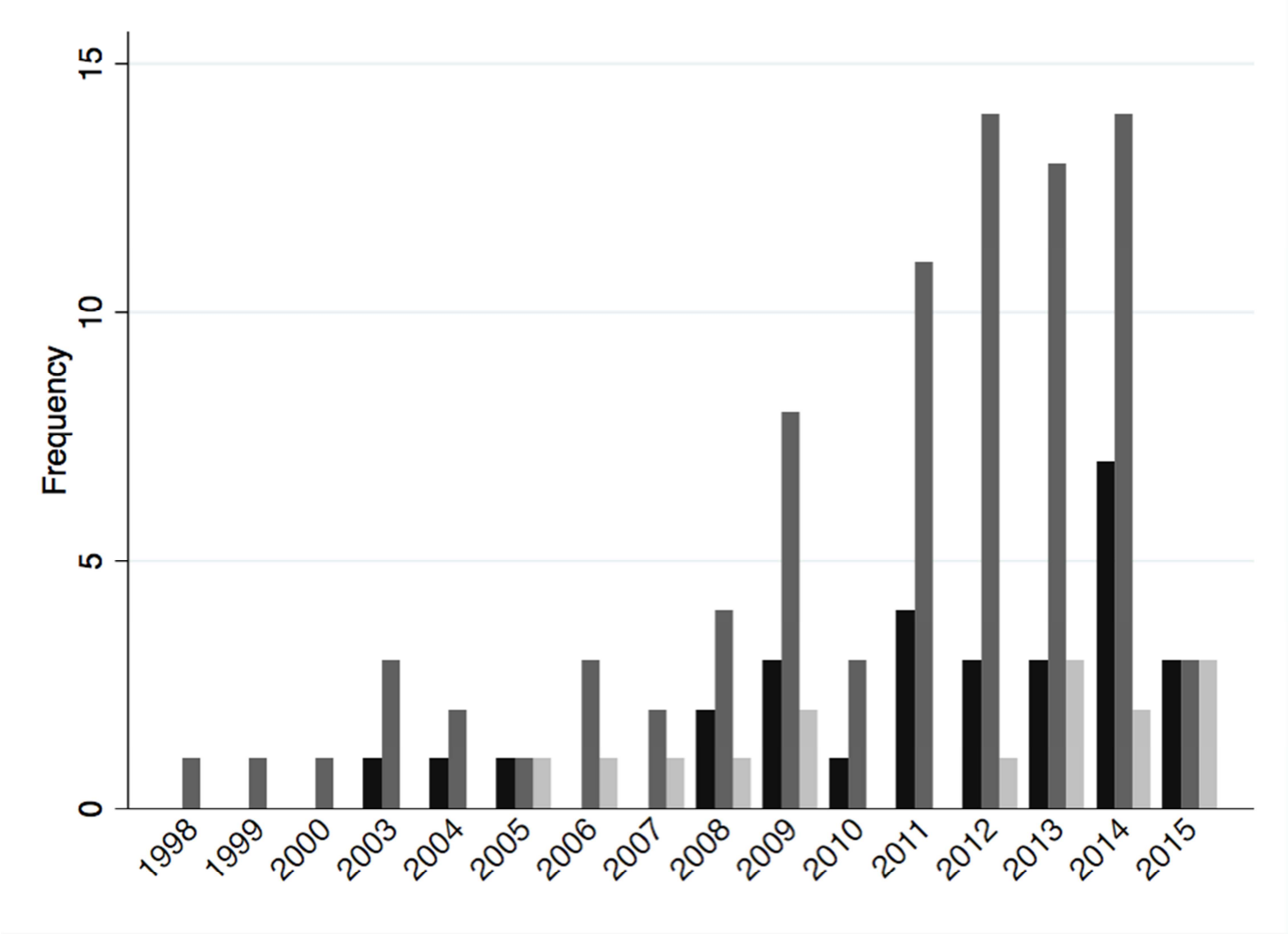


Figure 2: Plot of frequency of reviewed methods for mitigating for unmeasured confounding by: difference-in-differences [black]; Instrumental variable analysis (IVA) [mid-grey]; Other [light grey] includes regression discontinuity, prior event rate ratio method, propensity score calibration, perturbation analysis, negative control outcomes, fixed effects with IVA and dynamic panel models. Note: the low frequencies in 2015 was attributable to the May cut-off for inclusion in that year.

Tables

Method	Description	Obstacles to implementation	Frequency of methods
Instrumental variable analysis (IVA)	Upon identification of a suitably strong instrument, the influence of bias may be reduced through post-hoc randomisation. The instrumental variable should be highly determinant of the intervention or treatment received, while satisfying the exclusion assumption of being independent of the outcome other than through the treatment (Wright 1928; Angrist 1991).	In practice, finding an instrument with a sufficiently strong treatment association is a stumbling block in many analyses (Bound, Jaeger, and Baker 1995; Baser 2009). Association of the instrument with the outcome exclusively through the treatment is an untestable assumption, particularly if an indirect association exists through an unmeasured covariate.	79
Difference-in-differences (DiD)	A biased effect estimate between two treatment groups may be corrected by the same estimates from a treatment-free period prior to the exposure, which should be a measure of the confounding bias contributed to the treatment effect (Ashenfelter and Card 1984). Aggregated at the treatment group level, this is operationalised in regression as a period-treatment interaction. At an individual level, demeaning, first-differencing or dummy variables for each individual may yield bias-free fixed effects, contingent on assumptions.	The method is contingent on the availability of repeated outcomes in both periods and invokes a time-invariant confounding assumption: that the confounding bias as captured by the estimated treatment effect in a treatment-free period prior to exposure is constant through to the study period.	24
Prior event rate ratio (PERR)	Analogous to the DiD method for time-to-event or rate data, a biased estimate of the hazard ratio or the incidence rate ratio is adjusted through its ratio with that from a treatment-free prior period (Tannen et al. 2008).	As with the assumption for DiD, repeatable outcomes and a constancy of the unmeasured confounding bias is required across both periods, before and after the exposure. Prior event occurrence should not influence the likelihood of future treatment.	5
Fixed effects instrumental variable analysis (FE IVA)	IVA may be applied to DiD estimation to mitigate for second-order endogeneity: the time-varying part of the bias that may not have been adjusted for by DiD.	Assumptions of IVA apply	5
Dynamic panel model, or Instrumental variable - generalised method of moments (IV-GMM)	Lagged observations of the confounded (endogenous) explanatory variable are introduced in a first-differences fixed effects analysis so that the differences of the lags become the instrumental variables in a generalised method of moments estimation.	Assumptions of IVA apply. Here the differenced lags should not be correlated with the differences in the error terms.	2
Regression discontinuity (RD)	RD is a design for analysis based on a treatment assignment determined by a cut-off applied to a continuous variable that is preferably measured with some random noise (as many clinical tests may be). The outcome can then be modelled on treatment for individuals within a certain interval from the cut-off of the assignment variable to ensure exchangeability between individuals for robust causal inference (Thistlethwaite and Campbell 1960)	Where assignment is not sharply determined by the cut-off, an increase in the probability of treatment may be observed leading to a "fuzzy" version of RD. Continuity in the assignment variable is assumed, otherwise manipulation of assignment and reverse causality may be suspected. Assignment should be locally random around the cut-off and makes the weak assumption that no unobserved covariates are discontinuous around the assignment cut-off.	3
Propensity score calibration (PSC)	PSC adjusts for residual confounding in the error-prone main dataset by importing information about the unmeasured confounders from a smaller, external “gold-standard” dataset (Stürmer et al. 2005). Analysis in the main dataset is adjusted using a single dimension propensity score of the measured corrected for unmeasured confounding by regression calibration against the gold-standard propensity score.	Successful adjustment is wholly dependent on the availability of another dataset containing the exposure variable and error-free predictor, with individuals that are relevant enough to those in the main dataset and under similar enough conditions to assure sufficient overlap between the two datasets.	3
Perturbation testing/analysis (PT/PA)	This data mining approach aims to mitigate for unmeasured confounding by adjusting for many measured variables that are weakly associated with the unobserved confounding variables (Lee 2014). Simulation in the single reviewed example demonstrated this may require 100's, if not 1000's of perturbation variables (PV).	This requires a very highly dimensional dataset, which may ultimately obviate the need for indirect adjustment if the most or all of the confounders are captured. Simulation demonstrated the bias may be exaggerated if a confounder is inadvertently identified as a PV, requiring many more true PVs to correct the bias. The number of PVs may exceed the available degrees of freedom necessitating clustering.	1
Negative control outcome / exposure (NCO/NCE)	A negative controlis causally related to measured and unmeasured confounders affecting the exposure and main outcome, but not directly causally related to exposure and outcome themselves. As such, the negative control may be used to detect confounding bias in the main study, and potentially to indirectly adjust for this (Richardson et al. 2014)	This assumes that the effect of the unmeasured confounders on the main outcome is similar to that affecting the negative control.	1

Table 1: Summary of methods to mitigate against unmeasured confounding captured by systematic review, and the frequency of their use amongst the captured papers

IV type	Explanation/ Example	No. of papers				Total frequency
Mendelian	Genetic characteristics :Single nucleotide polymorphisms	11				11
Geographic	Differential distance between patient's postcode and nearest health facility	19	1			20
Time	Time-based characteristic of treatment such as date of therapy	6		2	1	10
Historical	Usually prescribing preference of physician or facility based on historical records of previously administered therapies	31				34
Lagged	Previous therapy or outcome of patient	6				6
Randomisation	Original randomisation	1				1
Other	Characteristics of individual e.g: age of patient, weight of offspring	8				8

Table 2: Frequency of instruments categorised by type used in instrumental variable analyses

Appendix A

1. ("prior event" and ratio).ti,ab.

2. "paired cox model".ti,ab.

3. 1 or 2

4. instrumental variables.ti,ab.

5. instrumental variable analysis/

6. propensity score calibration.ti,ab.

7. regression discontinuity design.ti,ab.

8. "difference in differences".ti,ab.

9. (difference adj1 differences).ti,ab.

10. "ratio of ratios".ti,ab.

11. (ratio adj1 ratios).ti,ab.

12. interrupted time series.ti,ab.

13. segmented regression.ti,ab.

14. (sensitivity analysis/ or sensitivity analysis.ti,ab.) and ((unmeasured or residual or hidden) and (confounding or confounder*)).ti,ab.

15. or/4-14

16. ((unmeasured or residual or hidden or unobserved or omitted) and (confounding or confounder*)).ti,ab.

17. confounding variable/

18. covariates.ti,ab.

19. bias.ti,ab.

20. selection bias/

21. 16 or 17 or 18 or 19 or 20

22. observational study/

23. (observation* adj (stud* or data)).ti,ab.

24. ((before adj after) and (study or studies)).ti,ab.

25. (nonrandomi?ed or non randomi?ed).ti,ab.

26. case crossover.ti,ab.

27. case control.ti,ab.

28. case control study/

29. cohort study.ti,ab.

30. (quasi experiment* or quasiexperiment*).ti,ab.

31. quasi-experimental study/

32. cross sectional study.ti,ab.

33. cross-sectional study/

34. simulation.ti,ab.

35. case time control.ti,ab.

36. ("before and after" and (study or studies)).ti,ab.

37. or/22-36

38. 16 and 19 and 37

39. 3 or 15

40. 39 and 37 and 21

41. 38 or 40

42. 21 or 37

43. 39 and 42

Appendix B

Table 3: Table of included studies denoting QE method used and type of instrument, if applicable, where: IVA = instrumental variable analysis; RD = regression discontinuity; DiD = difference-in-differences; DiDiD = difference-in-difference-in-differences; PSC = propensity score calibration; PERR = prior event rate ratio

Author	Title	Year	QE method	If IVA, IV type
Bryson, W. C.; McConnell, J.; Krothuis, T.; McCarty, D.	Extended-release naltrexone for alcohol dependence: persistence and healthcare costs and utilization	2011	DiD	
Cheng, L.; Liu, H.; Zhang, Y.; Shen, K.; Zeng, Y.	The impact of health insurance on health outcomes and spending of the elderly: Evidence from china's new cooperative medical scheme	2015	DiD	
Gebel, M.; Vosemer, J.	The impact of employment transitions on health in Germany. A difference-in-differences propensity score matching approach	2014	DiD	
Goetzel, R. Z.; Roemer, E. C.; Pei, X.; Short, M. E.; Tabrizi, M. J.; Wilson, M. G.; Dejoy, D. M.; Craun, B. A.; Tully, K. J.; White, J. M.; Baase, C. M.	Second-year results of an obesity prevention program at the dow chemical company	2010	DiD	
Higgins, S.; Chawla, R.; Colombo, C.; Snyder, R.; Nigam, S.	Medical homes and cost and utilization among high-risk patients	2014	DiD	
Kausto, J.; Viikari-Juntura, E.; Virta, L. J.; Gould, R.; Koskinen, A.; Solovieva, S.	Effectiveness of new legislation on partial sickness benefit on work participation: a quasi-experiment in Finland	2014	DiD	

Kelly, Y.; Kelly, J.; Sacker, A.	Changes in bedtime schedules and behavioral difficulties in 7 year old children	2013	DiD
Lin, W. C.; Chien, H. L.; Willis, G.; O'Connell, E.; Rennie, K. S.; Bottella, H. M.; Ferris, T. G.	The effect of a telephone-based health coaching disease management program on medicaid members with chronic conditions	2012	DiD
Lyon, S. M.; Wunsch, H.; Asch, D. A.; Carr, B. G.; Kahn, J. M.; Cooke, C. R.	Use of intensive care services and associated hospital mortality after massachusetts healthcare reform	2014	DiD
Menon, J.; Paulet, M.; Thomas, Iii J.	Wellness coaching and health-related quality of life: A case-control difference-in-differences analysis	2012	DiD
Moran, J. R.; Short, P. F.; Hollenbeak, C. S.	Long-term employment effects of surviving cancer	2011	DiD
Osborne, N. H.; Nicholas, L. H.; Ryan, A. M.; Thumma, J. R.; Dimick, J. B.	Association of hospital participation in a quality reporting program with surgical outcomes and expenditures for medicare beneficiaries	2015	DiD
Reid, R. O.; Ashwood, J. S.; Friedberg, M. W.; Weber, E. S.; Setodji, C. M.; Mehrotra, A.	Retail clinic visits and receipt of primary care	2013	DiD

Sadhu, A. R.; Ang, A. C.; Ingram-Drake, L. A.; Martinez, D. S.; Hsueh, W. A.; Ettner, S. L.	Economic benefits of intensive insulin therapy in critically Ill patients: The targeted insulin therapy to improve hospital outcomes (TRIUMPH) project	2008	DiD
Sarkar, U.; Lyles, C. R.; Parker, M. M.; Allen, J.; Nguyen, R.; Moffet, H. H.; Schillinger, D.; Karter, A. J.	Use of the refill function through an online patient portal is associated with improved adherence to statins in an integrated health system	2014	DiD
Watt, C.; Abuya, T.; Warren, C. E.; Obare, F.; Kanya, L.; Bellows, B.	Can reproductive health voucher programs improve quality of postnatal care? A quasi-experimental evaluation of Kenya ' s Safe Motherhood voucher scheme	2015	DiD
De Preux, L. B.	Anticipatory ex ante moral hazard and the effect of medicare on prevention	2011	DiD; DiDiD
Rajaram, R.; Chung, J. W.; Jones, A. T.; Cohen, M. E.; Dahlke, A. R.; Ko, C. Y.; Tarpley, J. L.; Lewis, F. R.; Hoyt, D. B.; Bilimoria, K. Y.	Association of the 2011 ACGME resident duty hour reform with general surgery patient outcomes and with resident examination performance	2014	DiD; DiDiD
Domino, M. E.; Norton, E. C.; Morrissey, J. P.; Thakur, N.	Cost shifting to jails after a change to managed mental health care	2004	DiD; Fixed effects
Hodgkin, D.; Parks Thomas, C.; Simoni-Wastila, L.; Ritter, G. A.; Lee, S.	The effect of a three-tier formulary on antidepressant utilization and expenditures	2008	Fixed effects

Li, J.; Hurley, J.; DeCicca, P.; Buckley, G.	Physician response to pay-for-performance: evidence from a natural experiment	2014	DiD pooled OLS; DiD (Fixed effects); DiD + differential trends	
Yoon, J.; Bernell, S. L.	The role of adverse physical health events on the utilization of mental health services	2013	DiD & Fixed Effects	
Fortney, J. C.; Steffick, D. E.; Burgess Jr, J. F.; Maciejewski, M. L.; Petersen, L. A.	Are primary care services a substitute or complement for specialty and inpatient services?	2005	IVA applied to DiD	Geographic
Hay, J.; Jhaveri, M.; Tangirala, M.; Kaliner, M.	Cost and resource utilization comparisons of second-generation antihistamines vs. montelukast for allergic rhinitis treatment	2009	IVA applied to Fixed effects	Historical
Chung, S.; Domino, M. E.; Stearns, S. C.	The effect of retirement on weight	2009	Fixed Effects; IVA applied to Fixed effects	Lagged
Wagner, T. H.; Jimison, H. B.	Computerized health information and the demand for medical care	2003	IVA applied to Fixed effects	Other
Kawatkar, A. A.; Hay, J. W.; Stohl, W.; Nichol, M. B.	Incremental expenditure of biologic disease modifying antirheumatic treatment using instrumental variables in panel data	2013	Dynamic panel model (IV-GMM)	Lagged

Piernas, C.; Ng, S. W.; Mendez, M. A.; Gordon-Larsen, P.; Popkin, B. M.	A dynamic panel model of the associations of sweetened beverage purchases with dietary quality and food-purchasing patterns	2015	Dynamic panel model (IV-GMM)	Lagged
Lei, X.; Lin, W.	The new cooperative medical scheme in rural China: Does more coverage mean more service and better health?	2009	Fixed effects; IVA; DiD	Geographic
Lin, M. J.; Liu, J. T.	Do lower birth weight babies have lower grades? Twin fixed effect and instrumental variable method evidence from Taiwan	2009	Fixed effects; IVA	Geographic
Schmittdiel, J. A.; Karter, A. J.; Dyer, W.; Parker, M.; Uratsu, C.; Chan, J.; Duru, O. K.	The comparative effectiveness of mail order pharmacy use vs. local pharmacy use on LDL-C control in new statin users	2011	DiD; IVA	Other
Basu, A.	Estimating Decision-Relevant Comparative Effects Using Instrumental Variables	2011	IVA	Geographic
Beck, C. A.; Penrod, J.; Gyorkos, T. W.; Shapiro, S.; Pilote, L.	Does Aggressive Care Following Acute Myocardial Infarction Reduce Mortality? Analysis with Instrumental Variables to Compare Effectiveness in Canadian and United States Patient Populations	2003	IVA	Geographic
Chen, L. F.; Chen, H. P.; Huang, Y. S.; Huang, K. Y.; Chou, P.; Lee, C. C.	Pneumococcal Pneumonia and the Risk of Stroke: A Population-Based Follow-Up Study	2012	IVA	Geographic

Edwards, S. T.; Prentice, J. C.; Simon, S. R.; Pizer, S. D.	Home-Based Primary Care and the risk of ambulatory care-sensitive condition hospitalization among older veterans with diabetes mellitus	2014	IVA	Geographic
Frances, C. D.; Shlipak, M. G.; Noguchi, H.; Heidenreich, P. A.; McClellan, M.	Does physician specialty affect the survival of elderly patients with myocardial infarction?	2000	IVA	Geographic
Goldman, D. P.; Bao, Y.	Effective HIV treatment and the employment of HIV+ adults	2004	IVA	Geographic
Gowrisankaran, G.; Town, R. J.	Estimating the quality of care in hospitals using instrumental variables	1999	IVA	Geographic
Hirth, R. A.; Grabowski, D. C.; Feng, Z.; Rahman, M.; Mor, V.	Effect of nursing home ownership on hospitalization of long-stay residents: An instrumental variables approach	2014	IVA	Geographic
Kahn, J. M.; Werner, R. M.; David, G.; Ten Have, T. R.; Benson, N. M.; Asch, D. A.	Effectiveness of long-term acute care hospitalization in elderly patients with chronic critical illness	2013	IVA	Geographic
Linden, A.; Adams, J. L.	Evaluating disease management programme effectiveness: An introduction to instrumental variables	2006	IVA	Geographic

Norton, E. C.; Lindrooth, R. C.; Ennett, S. T.	Controlling for the endogeneity of peer substance use on adolescent alcohol and tobacco use	1998	IVA	Geographic
Pilote, L.; Beck, C. A.; Eisenberg, M. J.; Humphries, K.; Joseph, L.; Penrod, J. R.; Tu, J. V.	Comparing invasive and noninvasive management strategies for acute myocardial infarction using administrative databases	2008	IVA	Geographic
Pracht, E. E.; Tepas, Iii J. J.; Celso, B. G.; Langland-Orban, B.; Flint, L.	Survival advantage associated with treatment of injury at designated trauma centers: A bivariate probit model with instrumental variables	2007	IVA	Geographic
Slade, E. P.; McCarthy, J. F.; Valenstein, M.; Visnic, S.; Dixon, L. B.	Cost savings from assertive community treatment services in an era of declining psychiatric inpatient use	2013	IVA	Geographic
Tsai, A. C.; Votruba, M.; Bridges, J. F. P.; Cebul, R. D.	Overcoming bias in estimating the volume-outcome relationship	2006	IVA	Geographic
Wehby, G. L.; Ullrich, F.; Xie, Y.	Very low birth weight hospital volume and mortality: An instrumental variables approach	2012	IVA	Geographic
Hadley, J.; Polsky, D.; Mandelblatt, J. S.; Mitchell, J. M.; Weeks, J. C.; Wang, Q.; Hwang, Y. T.	An exploratory instrumental variable analysis of the outcomes of localized breast cancer treatments in a medicare population	2003	IVA	Geographic + Historical + Time

O'Malley, A. J.; Frank, R. G.; Normand, S. L. T.	Estimating cost-offsets of new medications: Use of new antipsychotics and mental health costs for schizophrenia	2011	IVA	Geographic + Time
Abrahamowicz, M.; Beauchamp, M. E.; Ionescu-Ittu, R.; Delaney, J. A. C.; Pilote, L.	Reducing the variance of the prescribing preference-based instrumental variable estimates of the treatment effect	2011	IVA	Historical
An, J.; Nichol, M. B.	Multiple medication adherence and its effect on clinical outcomes among patients with comorbid type 2 diabetes and hypertension	2013	IVA	Historical
Bekelman, J. E.; Mitra, N.; Handorf, E. A.; Uzzo, R. G.; Hahn, S. A.; Polsky, D.; Armstrong, K.	Effectiveness of androgen-deprivation therapy and radiotherapy for older men with locally advanced prostate cancer	2015	IVA	Historical
Bhowmik, D.; Aparasu, R. R.; Rajan, S. S.; Sherer, J. T.; Ochoa-Perez, M.; Chen, H.	Risk of manic switch associated with antidepressant therapy in pediatric bipolar depression	2014	IVA	Historical
Brooks, J. M.; Tang, Y.; Chapman, C. G.; Cook, E. A.; Chrischilles, E. A.	What is the effect of area size when using local area practice style as an instrument?	2013	IVA	Historical
Chuang, C. M.; Chou, Y. J.; Yen, M. S.; Chao, K. C.; Twu, N. F.; Wu, H. H.; Wen, K. C.; Chen, Y. J.; Wang, P. H.; Lai, C. R.; Chou, P.	The role of secondary cytoreductive surgery in patients with recurrent epithelial ovarian, tubal, and peritoneal cancers: A comparative effectiveness analysis	2012	IVA	Historical

De Ridder, A.; De Graeve, D.	Can we account for selection bias? A comparison between bare metal and drug-eluting stents	2011	IVA	Historical
Fang, G.; Brooks, J. M.; Chrischilles, E. A.	Comparison of instrumental variable analysis using a new instrument with risk adjustment methods to reduce confounding by indication	2012	IVA	Historical
Figueroa, R.; Harman, J.; Engberg, J.	Use of Claims Data to Examine the Impact of Length of Inpatient Psychiatric Stay on Readmission Rate	2004	IVA	Historical
Huesch, M. D.	External adjustment sensitivity analysis for unmeasured confounding: An application to coronary stent outcomes, Pennsylvania 2004-2008	2013	IVA	Historical
Huybrechts, K. F.; Brookhart, M. A.; Rothman, K. J.; Silliman, R. A.; Gerhard, T.; Crystal, S.; Schneeweiss, S.	Comparison of different approaches to confounding adjustment in a study on the association of antipsychotic medication with mortality in older nursing home patients	2011	IVA	Historical
Ionescu-Ittu, R.	Treatment effect estimates varied depending on the definition of the provider prescribing preference-based instrumental variables	2012	IVA	Historical
Kivimaki, M.; Vahtera, J.; Kawachi, I.; Ferrie, J. E.; Oksanen, T.; Joensuu, M.; Pentti, J.; Salo, P.; Elovainio, M.; Virtanen, M.	Psychosocial work environment as a risk factor for absence with a psychiatric diagnosis: An instrumental-variables analysis	2010	IVA	Historical

Kramer, A.; Jager, K. J.; Fogarty, D. G.; Ravani, P.; Finne, P.; Perez-Panades, J.; Prutz, K. G.; Arias, M.; Heaf, J. G.; Wanner, C.; Stel, V. S.	Association between pre-transplant dialysis modality and patient and graft survival after kidney transplantation	2012	IVA	Historical
Kuo, Y. F.; Montie, J. E.; Shahinian, V. B.	Reducing bias in the assessment of treatment effectiveness: Androgen deprivation therapy for prostate cancer	2012	IVA	Historical
Lakdawalla, D. N.; Mascarenhas, M.; Jena, A. B.; Vanderpuye-Orgle, J.; Lavallee, C.; Linthicum, M. T.; Snider, J. T.	Impact of oral nutrition supplements on hospital outcomes in pediatric patients	2014	IVA	Historical
MacKenzie, T. A.; Tosteson, T. D.; Morden, N. E.; Stukel, T. A.; O'Malley, A. J.	Using instrumental variables to estimate a Cox's proportional hazards regression subject to additive confounding	2014	IVA	Historical
Margolis, D. J.; Gupta, J.; Hoffstad, O.; Papdopoulos, M.; Glick, H. A.; Thom, S. R.; Mitra, N.	Lack of effectiveness of hyperbaric oxygen therapy for the treatment of diabetic foot ulcer and the prevention of amputation a cohort study	2013	IVA	Historical
Parmar, A. D.; Sheffield, K. M.; Han, Y.; Vargas, G. M.; Guturu, P.; Kuo, Y. F.; Goodwin, J. S.; Riall, T. S.	Evaluating comparative effectiveness with observational data: Endoscopic ultrasound and survival in pancreatic cancer	2013	IVA	Historical
Pisoni, R. L.; Arrington, C. J.; Albert, J. M.; Ethier, J.; Kimata, N.; Krishnan, M.; Rayner, H. C.; Saito, A.; Sands, J. J.; Saran, R.; Gillespie, B.; Wolfe, R. A.; Port, F. K.	Facility Hemodialysis Vascular Access Use and Mortality in Countries Participating in DOPPS: An Instrumental Variable Analysis	2009	IVA	Historical

Prentice, J. C.; Conlin, P. R.; Gellad, W. F.; Edelman, D.; Lee, T. A.; Pizer, S. D.	Capitalizing on prescribing pattern variation to compare medications for type 2 diabetes	2014	IVA	Historical
Rassen, J. A.; Brookhart, M. A.; Glynn, R. J.; Mittleman, M. A.; Schneeweiss, S.	Instrumental variables II: instrumental variable application-in 25 variations, the physician prescribing preference generally was strong and reduced covariate imbalance	2009	IVA	Historical
Rosenthal, M. B.; Li, Z.; Robertson, A. D.; Milstein, A.	Impact of financial incentives for prenatal care on birth outcomes and spending	2009	IVA	Historical
Sheffield, K. M.; Riall, T. S.; Han, Y.; Kuo, Y. F.; Townsend, C. M., Jr.; Goodwin, J. S.	Association between cholecystectomy with vs without intraoperative cholangiography and risk of common duct injury	2013	IVA	Historical
Steingrub, J. S.; Lagu, T.; Rothberg, M. B.; Nathanson, B. H.; Raghunathan, K.; Lindenauer, P. K.	Treatment with neuromuscular blocking agents and the risk of in-hospital mortality among mechanically ventilated patients with severe sepsis	2014	IVA	Historical
Stukel, Thérèse A; Fisher, Elliott S; Wennberg, David E; Alter, David A; Gottlieb, Daniel J; Vermeulen, Marian J	Analysis of observational studies in the presence of treatment selection bias: effects of invasive cardiac management on AMI survival using propensity score and instrumental variable methods.	2007	IVA	Historical

Tagami, T.; Matsui, H.; Horiguchi, H.; Fushimi, K.; Yasunaga, H.	Antithrombin and mortality in severe pneumonia patients with sepsis-associated disseminated intravascular coagulation: An observational nationwide study	2014	IVA	Historical
VanDyke, R. D.; McPhail, G. L.; Huang, B.; Fenchel, M. C.; Amin, R. S.; Carle, A. C.; Chini, B. A.; Seid, M.	Inhaled tobramycin effectively reduces FEV1 decline in cystic fibrosis an instrumental variables analysis	2013	IVA	Historical
Wong, K.; Campitelli, M. A.; Stukel, T. A.; Kwong, J. C.	Estimating influenza vaccine effectiveness in community-dwelling elderly patients using the instrumental variable analysis method	2012	IVA	Historical
Chen, H.; Mehta, S.; Aparasu, R.; Patel, A.; Ochoa-Perez, M.	Comparative effectiveness of monotherapy with mood stabilizers versus second generation (atypical) antipsychotics for the treatment of bipolar disorder in children and adolescents	2014	IVA	Historical + Time
Newman, T. B.; Vittinghoff, E.; McCulloch, C. E.	Efficacy of phototherapy for newborns with hyperbilirubinemia: a cautionary example of an instrumental variable analysis	2012	IVA	Historical + Time
Ahern, T. P.; Pedersen, L.; Svaerke, C.; Rothman, K. J.; Sorensen, H. T.; Lash, T. L.	The association between vitamin K antagonist therapy and site-specific cancer incidence estimated by using heart valve replacement as an instrumental variable	2011	IVA	Lagged

Cai, B.; Hennessy, S.; Flory, J. H.; Sha, D.;Ten Have, T. R.; Small, D. S.	Simulation study of instrumental variable approaches with an application to a study of the antidiabetic effect of bezafibrate	2012	IVA	Lagged
O'Malley, A. J.	Instrumental variable specifications and assumptions for longitudinal analysis of mental health cost offsets	2012	IVA	Lagged
Cawley, J.; Meyerhoefer, C.	The medical care costs of obesity: An instrumental variables approach	2012	IVA	Other
Groenwold, R. H.; Hak, E.; Klungel, O. H.; Hoes, A. W.	Instrumental variables in influenza vaccination studies: mission impossible?!	2010	IVA	Other
Kim, D.; Leigh, J. P.	Estimating the effects of wages on obesity	2010	IVA	Other
Pirracchio, R.; Sprung, C.; Payen, D.; Chevret, S.	Benefits of ICU admission in critically ill patients: whether instrumental variable methods or propensity scores should be used	2011	IVA	Other
Selden, T. M.; Hudson, J. L.	Access to care and utilization among children: Estimating the effects of public and private coverage	2006	IVA	Other

Slade, E. P.; Wissow, L. S.; Davis, M.; Abrams, M. T.; Dixon, L. B.	Medicaid lapses and low-income young adults' receipt of outpatient mental health care after an inpatient stay	2014	IVA	Other
Hay, J. W.; Lawler, E.; Yucel, K.; Guo, A.; Balzer, T.; Gaziano, J. M.; Scranton, R. E.	Cost impact of diagnostic imaging for lower extremity peripheral vascular occlusive disease	2009	IVA	PScore (historical EHRs)
Guo, J.; Konetzka, R. T.; Manning, W. G.	The causal effects of home care use on institutional long-term care utilization and expenditures	2015	IVA	Randomisation
Federspiel, J. J.; Stearns, S. C.; Sheridan, B. C.; Kuritzky, J. J.; D'Arcy, L. P.; Crespin, D. J.; Carey, T. S.; Rossi, J. S.	Evaluating the effectiveness of a rapidly adopted cardiovascular technology with administrative data: The case of drug-eluting stents for acute coronary syndromes	2012	IVA	Time
Goyal, N.; Zubizarreta, J. R.; Small, D. S.; Lorch, S. A.	Length of stay and readmission among late preterm infants: An instrumental variable approach	2013	IVA	Time
Hollingsworth, J. M.; Norton, E. C.; Kaufman, S. R.; Smith, R. M.; Wolf Jr, J. S.; Hollenbeck, B. K.	Medical expulsive therapy versus early endoscopic stone removal for acute renal colic: An instrumental variable analysis	2013	IVA	Time
Johnston, K. M.; Gustafson, P.; Levy, A. R.; Grootendorst, P.	Use of instrumental variables in the analysis of generalized linear models in the presence of unmeasured confounding with applications to epidemiological research	2008	IVA	Time

O'Donnell, H. C.; Colman, G.; Trachtman, R. A.; Velazco, N.; Racine, A. D.	Impact of newborn follow-up visit timing on subsequent ED visits and hospital readmissions: AN instrumental variable analysis	2014	IVA	Time
Zeliadt, S. B.; Loggers, E. T.; Slatore, C. G.; Au, D. H.; Hebert, P. L.; Klein, G. J.; Kessler, L. G.; Backhus, L. M.	Preoperative PET and the reduction of unnecessary surgery among newly diagnosed lung cancer patients in a community setting	2014	IVA	Time
Brunner, E. J.; Kivimaki, M.; Witte, D. R.; Lawlor, D. A.; Davey Smith, G.; Cooper, J. A.; Miller, M.; Lowe, G. D.; Rumley, A.; Casas, J. P.; Shah, T.; Humphries, S. E.; Hingorani, A. D.; Marmot, M. G.; Timpson, N. J.; Kumari, M.	Inflammation, insulin resistance, and diabetes--Mendelian randomization using CRP haplotypes points upstream	2008	IVA (Mendelian)	Mendelian
Burgess, S.; Thompson, S. G.	Avoiding bias from weak instruments in mendelian randomization studies	2011	IVA (Mendelian)	Mendelian
Haring, R.; Teumer, A.; Volker, U.; Dorr, M.; Nauck, M.; Biffar, R.; Volzke, H.; Baumeister, S. E.; Wallaschofski, H.	Mendelian randomization suggests non-causal associations of testosterone with cardiometabolic risk factors and mortality	2013	IVA (Mendelian)	Mendelian
Jokela, M.; Elovainio, M.; Keltikangas-Jarvinen, L.; Batty, G. D.; Hintsanen, M.; Seppala, I.; Kahonen, M.; Viikari, J. S.; Raitakari, O. T.; Lehtimaki, T.; Kivimaki, M.	Body mass index and depressive symptoms: Instrumental-variables regression with genetic risk score	2012	IVA (Mendelian)	Mendelian

Kivimaki, M.; Magnussen, C. G.; Juonala, M.; Kahonen, M.; Kettunen, J.; Loo, B. M.; Lehtimaki, T.; Viikari, J.; Raitakari, O. T.	Conventional and Mendelian randomization analyses suggest no association between lipoprotein(a) and early atherosclerosis: The Young Finns Study	2011	IVA (Mendelian)	Mendelian
Laschkolnig, A.; Kollerits, B.; Lamina, C.; Meisinger, C.; Rantner, B.; Stadler, M.; Peters, A.; Koenig, W.; Stockl, A.; Dahnhardt, D.; Boger, C. A.; Kramer, B. K.; Fraedrich, G.; Strauch, K.; Kronenberg, F.	Lipoprotein (a) concentrations, apolipoprotein (a) phenotypes, and peripheral arterial disease in three independent cohorts	2014	IVA (Mendelian)	Mendelian
Lawlor, D. A.; Harbord, R. M.; Timpson, N. J.; Lowe, G. D.; Rumley, A.; Gaunt, T. R.; Baker, I.; Yarnell, J. W.; Kivimaki, M.; Kumari, M.; Norman, P. E.; Jamrozik, K.; Hankey, G. J.; Almeida, O. P.; Flicker, L.; Warrington, N.; Marmot, M. G.; Ben-Shlomo, Y.; Palmer, L. J.; Day, I. N.; Ebrahim, S.; Smith, G. D.	The association of C-reactive protein and CRP genotype with coronary heart disease: findings from five studies with 4,610 cases amongst 18,637 participants	2008	IVA (Mendelian)	Mendelian
Leong, A.; Rehman, W.; Dastani, Z.; Greenwood, C.; Timpson, N.; Langsetmo, L.; Berger, C.; Fu, L.; Wong, B. Y. L.; Malik, S.; Malik, R.; Hanley, D. A.; Cole, D. E. C.; Goltzman, D.; Richards, J. B.	The Causal Effect of Vitamin D Binding Protein (DBP) Levels on Calcemic and Cardiometabolic Diseases: A Mendelian Randomization Study	2014	IVA (Mendelian)	Mendelian

Nimptsch, K.; Aleksandrova, K.; Boeing, H.; Janke, J.; Lee, Y. A.; Jenab, M.; Bueno-De-Mesquita, H. B.; Jansen, E. H. J. M.; Tsilidis, K. K.; Trichopoulou, A.; Weiderpass, E.; Wu, C.; Overvad, K.; Tjonneland, A.; Boutron-Ruault, M. C.; Dossus, L.; Racine, A.; Kaaks, R.; Canzian, F.; Lagiou, P.; Trichopoulos, D.; Palli, D.; Agnoli, C.; Tumino, R.; Vineis, P.; Panico, S.; Johansson, A.; Van Guelpen, B.; Khaw, K. T.; Wareham, N.; Peeters, P. H.; Quiros, J. R.; Garcia, A. V.; Molina-Montes, E.; Dorronsoro, M.; Chirlaque, M. D.; Gurrea, A. B.; Key, T. J.; Duarte-Salles, T.; Stepien, M.; Gunter, M. J.; Riboli, E.; Pischon, T.	Association of CRP genetic variants with blood concentrations of C-reactive protein and colorectal cancer risk	2015	IVA (Mendelian)	Mendelian
Palmer, T. M.; Sterne, J. A. C.; Harbord, R. M.; Lawlor, D. A.; Sheehan, N. A.; Meng, S.; Granell, R.; Smith, G. D.; Didelez, V.	Instrumental variable estimation of causal risk ratios and causal odds ratios in mendelian randomization analyses	2011	IVA (Mendelian)	Mendelian
Wehby, G. L.; Scholder, Sv	Genetic instrumental variable studies of effects of prenatal risk factors	2013	IVA (Mendelian)	Mendelian
Richardson, D. B.; Laurier, D.; Schubauer-Berigan, M. K.; Tchetgen, E. T.; Cole, S. R.	Assessment and indirect adjustment for confounding by smoking in cohort studies using relative hazards models	2014	Negative Control Outcome	

Brophy, S.; Jones, K. H.; Rahman, M. A.; Zhou, S. M.; John, A.; Atkinson, M. D.; Francis, N.; Lyons, R. A.; Dunstan, F.	Incidence of campylobacter and salmonella infections following first prescription for PPI: A cohort study using routine data	2013	PERR
Tannen, R. L.	Use of primary care electronic medical record database in drug efficacy research on cardiovascular outcomes: Comparison of database and randomised controlled trial findings	2009	PERR
Tannen, R. L.; Weiner, M. G.; Xie, D.	Replicated studies of two randomized trials of angiotensin-converting enzyme inhibitors: Further empiric validation of the 'prior event rate ratio' to adjust for unmeasured confounding by indication	2008	PERR
Tannen, R.; Xie, D.; Wang, X.; Yu, M.; Weiner, M. G.	A new "Comparative Effectiveness" assessment strategy using the THIN database: Comparison of the cardiac complications of pioglitazone and rosiglitazone	2013	PERR
Uddin, M. J.; Groenwold, R. H. H.; Van Staa, T. P.; De Boer, A.; Belitser, S. V.; Hoes, A. W.; Roes, K. C. B.; Klungel, O. H.	Performance of prior event rate ratio adjustment method in pharmacoepidemiology: A simulation study	2015	PERR
Lee, W. C.	Detecting and correcting the bias of unmeasured factors using perturbation analysis: a data-mining approach	2014	Perturbation analysis

Lunt, M.; Glynn, R. J.; Rothman, K. J.; Avorn, J.; Sturmer, T.	Propensity score calibration in the absence of surrogacy	2012	PSC
Sturmer, T.	Performance of propensity score calibration - A simulation study	2007	PSC
Stürmer, Til, Schneeweiss, Sebastian, Avorn, Jerry;	Adjusting effect estimates for unmeasured confounding with validation data using propensity score calibration	2005	PSC
Albouy, V.; Lequien, L.	Does compulsory education lower mortality?	2009	RD
Swaminathan, S.; Mor, V.; Mehrotra, R.; Trivedi, A. N.	Effect of medicare dialysis payment reform on use of erythropoiesis stimulating agents	2015	RD
Zuckerman, I. H.; Lee, E.; Wutoh, A. K.; Xue, Z.; Stuart, B.	Application of regression-discontinuity analysis in pharmaceutical health services research	2006	RD

References (included studies)

7. Stürmer T, Schneeweiss S, Avorn J, Glynn RJ. Adjusting effect estimates for unmeasured confounding with validation data using propensity score calibration. *Am J Epidemiol*. 2005;162(3):279-289. doi:10.1093/aje/kwi192.
8. Stürmer T, Schneeweiss S, Rothman KJ, Avorn J, Glynn RJ. Performance of propensity score calibration - A simulation study. *Am J Epidemiol*. 2007;165(10):1110-1118. doi:http://dx.doi.org/10.1093/aje/kwm074.
9. Lunt M, Glynn RJ, Rothman KJ, Avorn J, Stürmer T. Propensity score calibration in the absence of surrogacy. *Am J Epidemiol*. 2012;175(12):1294-1302. doi:10.1093/aje/kwr463.
10. Lee W-C. Detecting and correcting the bias of unmeasured factors using perturbation analysis: a data-mining approach. *BMC Med Res Methodol*. 2014;14(1):18. doi:10.1186/1471-2288-14-18.
11. Brunner EJ, Kivimaki M, Witte DR, et al. Inflammation, insulin resistance, and diabetes--Mendelian randomization using CRP haplotypes points upstream. *PLoS Med*. 2008;5(8):e155. internal-pdf://137.108.247.40/Infl_diabet_MendRand-Brunner-PLoS08.pdf.
12. Burgess S, Thompson SG. Avoiding bias from weak instruments in mendelian randomization studies. *Int J Epidemiol*. 2011;40(3):755-764. doi:http://dx.doi.org/10.1093/ije/dyr036.
13. Haring R, Teumer A, Völker U, et al. Mendelian randomization suggests non-causal associations of testosterone with cardiometabolic risk factors and mortality. *Andrology*. 2013;1(1):17-23. doi:10.1111/j.2047-2927.2012.00002.x.
14. Jokela M, Elovainio M, Keltikangas-Jarvinen L, et al. Body mass index and depressive symptoms: Instrumental-variables regression with genetic risk score. *Genes, Brain Behav*. 2012;11(8):942-948. doi:http://dx.doi.org/10.1111/j.1601-183X.2012.00846.x.
15. Kivimaki M, Magnussen CG, Juonala M, et al. Conventional and Mendelian randomization analyses suggest no association between lipoprotein(a) and early atherosclerosis: The Young Finns Study. *Int J Epidemiol*. 2011;40(2):470-478. doi:http://dx.doi.org/10.1093/ije/dyq205.
16. Laschkolnig A, Kollerits B, Lamina C, et al. Lipoprotein (a) concentrations, apolipoprotein (a) phenotypes, and peripheral arterial disease in three independent cohorts. *Cardiovasc Res*. 2014;103(1):28-36. doi:http://dx.doi.org/10.1093/cvr/cvu107.
17. Lawlor DA, Harbord RM, Timpson NJ, et al. The association of C-reactive protein and CRP genotype with coronary heart disease: findings from five studies with 4,610 cases amongst 18,637 participants. *PLoS ONE [Electronic Resour*. 2008;3(8):e3011. doi:http://dx.doi.org/10.1371/journal.pone.0003011.
18. Leong A, Rehman W, Dastani Z, et al. The Causal Effect of Vitamin D Binding Protein (DBP) Levels on Calcemic and Cardiometabolic Diseases: A Mendelian Randomization Study. *PLoS Med*. 2014;11(10):e1001751. doi:http://dx.doi.org/10.1371/journal.pmed.1001751.
19. Nimptsch K, Aleksandrova K, Boeing H, et al. Association of CRP genetic variants with blood concentrations of C-reactive protein and colorectal cancer risk. *Int J Cancer*. 2015;136(5):1181-1192. doi:http://dx.doi.org/10.1002/ijc.29086.

20. Palmer TM, Sterne JAC, Harbord RM, et al. Instrumental variable estimation of causal risk ratios and causal odds ratios in mendelian randomization analyses. *Am J Epidemiol*. 2011;173(12):1392-1403. doi:http://dx.doi.org/10.1093/aje/kwr026.
21. Wehby GL, Scholder S. Genetic instrumental variable studies of effects of prenatal risk factors. *Biodemography Soc Biol*. 2013;59(1):4-36. doi:http://dx.doi.org/10.1080/19485565.2013.774615.
22. Basu A. Estimating Decision-Relevant Comparative Effects Using Instrumental Variables. *Stat Biosci*. 2011;3(1):6-27. doi:http://dx.doi.org/10.1007/s12561-011-9033-6.
23. Beck CA, Penrod J, Gyorkos TW, Shapiro S, Pilote L. Does Aggressive Care Following Acute Myocardial Infarction Reduce Mortality? Analysis with Instrumental Variables to Compare Effectiveness in Canadian and United States Patient Populations. *Health Serv Res*. 2003;38(6):1423-1440. internal-pdf://133.21.114.118/MI_mort_IV-Beck-HSR03.pdf.
24. Cawley J, Meyerhoefer C. The medical care costs of obesity: An instrumental variables approach. *J Health Econ*. 2012;31:219-230. doi:http://dx.doi.org/10.1016/j.jhealeco.2011.10.003.
25. Chen LF, Chen HP, Huang YS, Huang KY, Chou P, Lee CC. Pneumococcal Pneumonia and the Risk of Stroke: A Population-Based Follow-Up Study. *PLoS One*. 2012;7(12):e51452. doi:http://dx.doi.org/10.1371/journal.pone.0051452.
26. Edwards ST, Prentice JC, Simon SR, Pizer SD. Home-Based Primary Care and the risk of ambulatory care-sensitive condition hospitalization among older veterans with diabetes mellitus. *JAMA Intern Med*. 2014;174(11):1796-1803. doi:http://dx.doi.org/10.1001/jamainternmed.2014.4327.
27. Federspiel JJ, Stearns SC, Sheridan BC, et al. Evaluating the effectiveness of a rapidly adopted cardiovascular technology with administrative data: The case of drug-eluting stents for acute coronary syndromes. *Am Heart J*. 2012;164(2):207-214. doi:http://dx.doi.org/10.1016/j.ahj.2012.05.016.
28. Frances CD, Shlipak MG, Noguchi H, Heidenreich PA, McClellan M. Does physician specialty affect the survival of elderly patients with myocardial infarction? *Health Serv Res*. 2000;35(5):1093-1116. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed5&AN=2001006754.
29. Goldman DP, Bao Y. Effective HIV treatment and the employment of HIV+ adults. *Health Serv Res*. 2004;39(6):1691-1712. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed6&AN=2004522727; http://lib.exeter.ac.uk:4556/resserv?sid=OVID:embase&id=pmid:&id=doi:&issn=0017-9124&isbn=&volume=39&issue=6+I&spage=1691&pages=1691-1712&date=2004&title=Health+Serv.
30. Gowrisankaran G, Town RJ. Estimating the quality of care in hospitals using instrumental variables. *J Health Econ*. 1999;18(6):747-767. doi:http://dx.doi.org/10.1016/S0167-6296%2899%2900022-3.
31. Goyal N, Zubizarreta JR, Small DS, Lorch SA. Length of stay and readmission among late preterm infants: An instrumental variable approach. *Hosp Pediatr*. 2013;3(1):7-15. doi:http://dx.doi.org/10.1542/hpeds.2012-0027.
32. Groenwold RH, Hak E, Klungel OH, Hoes AW. Instrumental variables in influenza vaccination studies: mission impossible?! *Value Heal*. 2010;13(1):132-137.

doi:http://dx.doi.org/10.1111/j.1524-4733.2009.00584.x.
33. Hirth RA, Grabowski DC, Feng Z, Rahman M, Mor V. Effect of nursing home ownership on hospitalization of long-stay residents: An instrumental variables approach. *Int J Health Care Finance Econ*. 2014;14(1):1-18. doi:http://dx.doi.org/10.1007/s10754-013-9136-3.
34. Hollingsworth JM, Norton EC, Kaufman SR, Smith RM, Wolf Jr JS, Hollenbeck BK. Medical expulsive therapy versus early endoscopic stone removal for acute renal colic: An instrumental variable analysis. *J Urol*. 2013;190(3):882-887. doi:http://dx.doi.org/10.1016/j.juro.2013.03.040.
35. Johnston KM, Gustafson P, Levy AR, Grootendorst P. Use of instrumental variables in the analysis of generalized linear models in the presence of unmeasured confounding with applications to epidemiological research. *Stat Med*. 2008;27(9):1539-1556. doi:10.1002/sim.3036.
36. Kahn JM, Werner RM, David G, Ten Have TR, Benson NM, Asch DA. Effectiveness of long-term acute care hospitalization in elderly patients with chronic critical illness. *Med Care*. 2013;51(1):4-10. doi:http://dx.doi.org/10.1097/MLR.0b013e31826528a7.
37. Kim D, Leigh JP. Estimating the effects of wages on obesity. *J Occup Environ Med*. 2010;52(5):495-500. doi:http://dx.doi.org/10.1097/JOM.0b013e3181dbc867.
38. Lei X, Lin W. The new cooperative medical scheme in rural China: Does more coverage mean more service and better health? *Health Econ*. 2009;18(suppl. 2):S25-S46. doi:http://dx.doi.org/10.1002/hec.1501.
39. Lin MJ, Liu JT. Do lower birth weight babies have lower grades? Twin fixed effect and instrumental variable method evidence from Taiwan. *Soc Sci Med*. 2009;68(10):1780-1787. doi:http://dx.doi.org/10.1016/j.socscimed.2009.02.031.
40. Linden A, Adams JL. Evaluating disease management programme effectiveness: An introduction to instrumental variables. *J Eval Clin Pract*. 2006;12(2):148-154. doi:http://dx.doi.org/10.1111/j.1365-2753.2006.00615.x.
41. Norton EC, Lindrooth RC, Ennett ST. Controlling for the endogeneity of peer substance use on adolescent alcohol and tobacco use. *Health Econ*. 1998;7(5):439-453. doi:http://dx.doi.org/10.1002/%28SICI%291099-1050%28199808%297:5%3C439::AID-HEC362%3E3.0.CO;2-9.
42. O'Donnell HC, Colman G, Trachtman RA, Velazco N, Racine AD. Impact of newborn follow-up visit timing on subsequent ED visits and hospital readmissions: AN instrumental variable analysis. *Acad Pediatr*. 2014;14(1):84-91. doi:http://dx.doi.org/10.1016/j.acap.2013.09.010.
43. O'Malley AJ, Frank RG, Normand SLT. Estimating cost-offsets of new medications: Use of new antipsychotics and mental health costs for schizophrenia. *Stat Med*. 2011;30(16):1971-1988. doi:http://dx.doi.org/10.1002/sim.4245.
44. Pilote L, Beck CA, Eisenberg MJ, et al. Comparing invasive and noninvasive management strategies for acute myocardial infarction using administrative databases. *Am Heart J*. 2008;155(1):42-48. doi:http://dx.doi.org/10.1016/j.ahj.2007.09.016.
45. Pirracchio R, Sprung C, Payen D, Chevret S. Benefits of ICU admission in critically ill patients: whether instrumental variable methods or propensity scores should be used. *BMC Med Res Methodol*. 2011;11(132). http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed11&AN=21936926.

46. Pracht EE, Tepas IJJ, Celso BG, Langland-Orban B, Flint L. Survival advantage associated with treatment of injury at designated trauma centers: A bivariate probit model with instrumental variables. *Med Care Res Rev*. 2007;64(1):83-97. doi:http://dx.doi.org/10.1177/1077558706296241.
47. Schmittdiel JA, Karter AJ, Dyer W, et al. The comparative effectiveness of mail order pharmacy use vs. local pharmacy use on LDL-C control in new statin users. *J Gen Intern Med*. 2011;26(12):1396-1402. doi:http://dx.doi.org/10.1007/s11606-011-1805-7.
48. Selden TM, Hudson JL. Access to care and utilization among children: Estimating the effects of public and private coverage. *Med Care*. 2006;44(5 SUPPL.):I19-I26. doi:http://dx.doi.org/10.1097/01.mlr.0000208137.46917.3b.
49. Slade EP, McCarthy JF, Valenstein M, Visnic S, Dixon LB. Cost savings from assertive community treatment services in an era of declining psychiatric inpatient use. *Health Serv Res*. 2013;48(1):195-217. doi:http://dx.doi.org/10.1111/j.1475-6773.2012.01420.x.
50. Slade EP, Wissow LS, Davis M, Abrams MT, Dixon LB. Medicaid lapses and low-income young adults' receipt of outpatient mental health care after an inpatient stay. *Psychiatr Serv*. 2014;65(4):454-460. doi:http://dx.doi.org/10.1176/appi.ps.201200375.
51. Tsai AC, Votruba M, Bridges JFP, Cebul RD. Overcoming bias in estimating the volume-outcome relationship. *Health Serv Res*. 2006;41(1):252-264. doi:http://dx.doi.org/10.1111/j.1475-6773.2005.00461.x.
52. Wehby GL, Ullrich F, Xie Y. Very low birth weight hospital volume and mortality: An instrumental variables approach. *Med Care*. 2012;50(8):714-721. doi:http://dx.doi.org/10.1097/MLR.0b013e31824e32cf.
53. Zeliadt SB, Loggers ET, Slatore CG, et al. Preoperative PET and the reduction of unnecessary surgery among newly diagnosed lung cancer patients in a community setting. *J Nucl Med*. 2014;55(3):379-385. doi:http://dx.doi.org/10.2967/jnumed.113.124230.
54. O'Malley AJ. Instrumental variable specifications and assumptions for longitudinal analysis of mental health cost offsets. *Heal Serv Outcomes Res Methodol*. 2012;12(4):254-272. doi:http://dx.doi.org/10.1007/s10742-012-0097-7.
55. Richardson DB, Laurier D, Schubauer-Berigan MK, Tchetgen ET, Cole SR. Assessment and indirect adjustment for confounding by smoking in cohort studies using relative hazards models. *Am J Epidemiol*. 2014;180(9):933-940. doi:http://dx.doi.org/10.1093/aje/kwu211.
57. Abrahamowicz M, Beauchamp ME, Ionescu-Ittu R, Delaney JAC, Pilote L. Reducing the variance of the prescribing preference-based instrumental variable estimates of the treatment effect. *Am J Epidemiol*. 2011;174(4):494-502. doi:http://dx.doi.org/10.1093/aje/kwr057.
58. Ahern TP, Pedersen L, Svaerke C, Rothman KJ, Sorensen HT, Lash TL. The association between vitamin K antagonist therapy and site-specific cancer incidence estimated by using heart valve replacement as an instrumental variable. *Am J Epidemiol*. 2011;174(12):1382-1390. doi:http://dx.doi.org/10.1093/aje/kwr268.
59. An J, Nichol MB. Multiple medication adherence and its effect on clinical outcomes among patients with comorbid type 2 diabetes and hypertension. *Med Care*. 2013;51(10):879-887. doi:http://dx.doi.org/10.1097/MLR.0b013e31829fa8ed.

60. Bekelman JE, Mitra N, Handorf EA, et al. Effectiveness of androgen-deprivation therapy and radiotherapy for older men with locally advanced prostate cancer. *J Clin Oncol*. 2015;33(7):716-722. doi:http://dx.doi.org/10.1200/JCO.2014.57.2743.
61. Bhowmik D, Aparasu RR, Rajan SS, Sherer JT, Ochoa-Perez M, Chen H. Risk of manic switch associated with antidepressant therapy in pediatric bipolar depression. *J Child Adolesc Psychopharmacol*. 2014;24(10):551-561. doi:http://dx.doi.org/10.1089/cap.2014.0028.
62. Brooks JM, Tang Y, Chapman CG, Cook EA, Chrischilles EA. What is the effect of area size when using local area practice style as an instrument? *J Clin Epidemiol*. 2013;66(8 SUPPL.8):S69-S83. doi:http://dx.doi.org/10.1016/j.jclinepi.2013.04.008.
63. Cai B, Hennessy S, Flory JH, Sha D, Ten Have TR, Small DS. Simulation study of instrumental variable approaches with an application to a study of the antidiabetic effect of bezafibrate. *Pharmacoepidemiol Drug Saf*. 2012;21(SUPPL.2):114-120. doi:http://dx.doi.org/10.1002/pds.3252.
64. Chen H, Mehta S, Aparasu R, Patel A, Ochoa-Perez M. Comparative effectiveness of monotherapy with mood stabilizers versus second generation (atypical) antipsychotics for the treatment of bipolar disorder in children and adolescents. *Pharmacoepidemiol Drug Saf*. 2014;23(3):299-308. doi:http://dx.doi.org/10.1002/pds.3568.
65. Chuang CM, Chou YJ, Yen MS, et al. The role of secondary cytoreductive surgery in patients with recurrent epithelial ovarian, tubal, and peritoneal cancers: A comparative effectiveness analysis. *Oncologist*. 2012;17(6):847-855. doi:http://dx.doi.org/10.1634/theoncologist.2011-0373.
66. De Ridder A, De Graeve D. Can we account for selection bias? A comparison between bare metal and drug-eluting stents. *Value Heal*. 2011;14(1):3-14. doi:http://dx.doi.org/10.1016/j.jval.2010.10.014.
67. Fang G, Brooks JM, Chrischilles EA. Comparison of instrumental variable analysis using a new instrument with risk adjustment methods to reduce confounding by indication. *Am J Epidemiol*. 2012;175(11):1142-1151. doi:http://dx.doi.org/10.1093/aje/kwr448.
68. Figueroa R, Harman J, Engberg J. Use of Claims Data to Examine the Impact of Length of Inpatient Psychiatric Stay on Readmission Rate. *Psychiatr Serv*. 2004;55(5):560-565. doi:http://dx.doi.org/10.1176/appi.ps.55.5.560.
69. Guo J, Konetzka RT, Manning WG. The causal effects of home care use on institutional long-term care utilization and expenditures. *Heal Econ (United Kingdom)*. 2015;24(S1):4-17. doi:http://dx.doi.org/10.1002/hec.3155.
70. Hadley J, Polsky D, Mandelblatt JS, et al. An exploratory instrumental variable analysis of the outcomes of localized breast cancer treatments in a medicare population. *Health Econ*. 2003;12(3):171-186. doi:http://dx.doi.org/10.1002/hec.710.
71. Hay JW, Lawler E, Yucel K, et al. Cost impact of diagnostic imaging for lower extremity peripheral vascular occlusive disease. *Value Heal*. 2009;12(2):262-266. doi:http://dx.doi.org/10.1111/j.1524-4733.2008.00438.x.
72. Huesch MD. External adjustment sensitivity analysis for unmeasured confounding: An application to coronary stent outcomes, Pennsylvania 2004-2008. *Health Serv Res*. 2013;48(3):1191-1214. doi:http://dx.doi.org/10.1111/1475-6773.12013.

73. Huybrechts KF, Brookhart MA, Rothman KJ, et al. Comparison of different approaches to confounding adjustment in a study on the association of antipsychotic medication with mortality in older nursing home patients. *Am J Epidemiol*. 2011;174(9):1089-1099. doi:http://dx.doi.org/10.1093/aje/kwr213.
74. Ionescu-Ittu R, Abrahamowicz M, Pilote L. Treatment effect estimates varied depending on the definition of the provider prescribing preference-based instrumental variables. *J Clin Epidemiol*. 2012;65(2):155-162. doi:http://dx.doi.org/10.1016/j.jclinepi.2011.06.012.
75. Kivimaki M, Vahtera J, Kawachi I, et al. Psychosocial work environment as a risk factor for absence with a psychiatric diagnosis: An instrumental-variables analysis. *Am J Epidemiol*. 2010;172(2):167-172. doi:http://dx.doi.org/10.1093/aje/kwq094.
76. Kramer A, Jager KJ, Fogarty DG, et al. Association between pre-transplant dialysis modality and patient and graft survival after kidney transplantation. *Nephrol Dial Transplant*. 2012;27(12):4473-4480. doi:http://dx.doi.org/10.1093/ndt/gfs450.
77. Kuo YF, Montie JE, Shahinian VB. Reducing bias in the assessment of treatment effectiveness: Androgen deprivation therapy for prostate cancer. *Med Care*. 2012;50(5):374-380. doi:http://dx.doi.org/10.1097/MLR.0b013e318245a086.
78. Lakdawalla DN, Mascarenhas M, Jena AB, et al. Impact of oral nutrition supplements on hospital outcomes in pediatric patients. *J Parenter Enter Nutr*. 2014;38(6):42S-49S. doi:http://dx.doi.org/10.1177/0148607114549769.
79. MacKenzie TA, Tosteson TD, Morden NE, Stukel TA, O'Malley AJ. Using instrumental variables to estimate a Cox's proportional hazards regression subject to additive confounding. *Heal Serv Outcomes Res Methodol*. 2014;14(1-2):54-68. doi:http://dx.doi.org/10.1007/s10742-014-0117-x.
80. Margolis DJ, Gupta J, Hoffstad O, et al. Lack of effectiveness of hyperbaric oxygen therapy for the treatment of diabetic foot ulcer and the prevention of amputation a cohort study. *Diabetes Care*. 2013;36(7):1961-1966. doi:http://dx.doi.org/10.2337/dc12-2160.
81. Newman TB, Vittinghoff E, McCulloch CE. Efficacy of phototherapy for newborns with hyperbilirubinemia: a cautionary example of an instrumental variable analysis. *Med Decis Making*. 2012;32(1):83-92. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed10&AN=21859678.
82. Parmar AD, Sheffield KM, Han Y, et al. Evaluating comparative effectiveness with observational data: Endoscopic ultrasound and survival in pancreatic cancer. *Cancer*. 2013;119(21):3861-3869. doi:http://dx.doi.org/10.1002/cncr.28295.
83. Pisoni RL, Arrington CJ, Albert JM, et al. Facility Hemodialysis Vascular Access Use and Mortality in Countries Participating in DOPPS: An Instrumental Variable Analysis. *Am J Kidney Dis*. 2009;53(3):475-491. doi:http://dx.doi.org/10.1053/j.ajkd.2008.10.043.
84. Prentice JC, Conlin PR, Gellad WF, Edelman D, Lee TA, Pizer SD. Capitalizing on prescribing pattern variation to compare medications for type 2 diabetes. *Value Heal*. 2014;17(8):854-862. doi:http://dx.doi.org/10.1016/j.jval.2014.08.2674.
85. Rassen JA, Brookhart MA, Glynn RJ, Mittleman MA, Schneeweiss S. Instrumental variables II: instrumental variable application-in 25 variations, the physician prescribing preference generally was strong and reduced covariate imbalance. *J Clin Epidemiol*. 2009;62(12):1233-1241. doi:http://dx.doi.org/10.1016/j.jclinepi.2008.12.006.

86. Rosenthal MB, Li Z, Robertson AD, Milstein A. Impact of financial incentives for prenatal care on birth outcomes and spending. *Health Serv Res*. 2009;44(5 Part 1):1465-1479. doi:http://dx.doi.org/10.1111/j.1475-6773.2009.00996.x.
87. Sheffield KM, Riall TS, Han Y, Kuo YF, Townsend, C. M. J, Goodwin JS. Association between cholecystectomy with vs without intraoperative cholangiography and risk of common duct injury. *Jama*. 2013;310(8):812-820. doi:http://dx.doi.org/10.1001/jama.2013.276205.
88. Steingrub JS, Lagu T, Rothberg MB, Nathanson BH, Raghunathan K, Lindenauer PK. Treatment with neuromuscular blocking agents and the risk of in-hospital mortality among mechanically ventilated patients with severe sepsis. *Crit Care Med*. 2014;42(1):90-96. doi:http://dx.doi.org/10.1097/CCM.0b013e31829eb7c9.
89. Stukel TA, Fisher ES, Wennberg DE, Alter DA, Gottlieb DJ, Vermeulen MJ. Analysis of observational studies in the presence of treatment selection bias: effects of invasive cardiac management on AMI survival using propensity score and instrumental variable methods. *JAMA*. 2007;297(3):278-285. doi:10.1001/jama.297.3.278.
90. Tagami T, Matsui H, Horiguchi H, Fushimi K, Yasunaga H. Antithrombin and mortality in severe pneumonia patients with sepsis-associated disseminated intravascular coagulation: An observational nationwide study. *J Thromb Haemost*. 2014;12(9):1470-1479. doi:http://dx.doi.org/10.1111/jth.12643.
91. VanDyke RD, McPhail GL, Huang B, et al. Inhaled tobramycin effectively reduces FEV1 decline in cystic fibrosis an instrumental variables analysis. *Ann Am Thorac Soc*. 2013;10(3):205-212. doi:http://dx.doi.org/10.1513/AnnalsATS.201209-082OC.
92. Wong K, Campitelli MA, Stukel TA, Kwong JC. Estimating influenza vaccine effectiveness in community-dwelling elderly patients using the instrumental variable analysis method. *Arch Intern Med*. 2012;172(6):484-491. doi:10.1001/archinternmed.2011.2038.
93. Bryson WC, McConnell J, Krothuis T, McCarty D. Extended-release naltrexone for alcohol dependence: persistence and healthcare costs and utilization. *Am J Manag Care*. 2011;17 Suppl 8:S222-234. internal-pdf://72.27.23.57/alcohol_depend-Bryson-AmJManCare11.pdf internal-pdf://4105231440/Bryson2011-suppl_appdx.docx.
94. Cheng L, Liu H, Zhang Y, Shen K, Zeng Y. The impact of health insurance on health outcomes and spending of the elderly: Evidence from china's new cooperative medical scheme. *Heal Econ (United Kingdom)*. 2015;24(6):672-691. doi:http://dx.doi.org/10.1002/hec.3053.
95. De Preux LB. Anticipatory ex ante moral hazard and the effect of medicare on prevention. *Health Econ*. 2011;20(9):1056-1072. doi:http://dx.doi.org/10.1002/hec.1778.
96. Gebel M, Voßemer J. The impact of employment transitions on health in Germany. A difference-in-differences propensity score matching approach. *Soc Sci Med*. 2014;108:128-136. doi:10.1016/j.socscimed.2014.02.039.
97. Goetzel RZ, Roemer EC, Pei X, et al. Second-year results of an obesity prevention program at the dow chemical company. *J Occup Environ Med*. 2010;52(3):291-302. doi:http://dx.doi.org/10.1097/JOM.0b013e3181d46f0b.
98. Higgins S, Chawla R, Colombo C, Snyder R, Nigam S. Medical homes and cost and utilization among high-risk patients. *Am J Manag Care*. 2014;20(3):e61-e71. internal-pdf://67.10.22.98/medhomes_cost_highrisk-Higgins-AJMC14.pdf.

99. Kausto J, Viikari-Juntura E, Virta LJ, Gould R, Koskinen A, Solovieva S. Effectiveness of new legislation on partial sickness benefit on work participation: a quasi-experiment in Finland. *BMJ Open*. 2014;4(12):e006685. doi:http://dx.doi.org/10.1136/bmjopen-2014-006685.
100. Kelly Y, Kelly J, Sacker A. Changes in bedtime schedules and behavioral difficulties in 7 year old children. *Pediatrics*. 2013;132(5):e1184-e1193. doi:http://dx.doi.org/10.1542/peds.2013-1906.
101. Lin WC, Chien HL, Willis G, et al. The effect of a telephone-based health coaching disease management program on medicaid members with chronic conditions. *Med Care*. 2012;50(1):91-98. doi:http://dx.doi.org/10.1097/MLR.0b013e31822dcedf.
102. Lyon SM, Wunsch H, Asch DA, Carr BG, Kahn JM, Cooke CR. Use of intensive care services and associated hospital mortality after massachusetts healthcare reform. *Crit Care Med*. 2014;42(4):763-770. doi:http://dx.doi.org/10.1097/CCM.0000000000000044.
103. Menon J, Paulet M, Thomas IJ. Wellness coaching and health-related quality of life: A case-control difference-in-differences analysis. *J Occup Environ Med*. 2012;54(10):1259-1267. doi:http://dx.doi.org/10.1097/JOM.0b013e31825a2594.
104. Moran JR, Short PF, Hollenbeak CS. Long-term employment effects of surviving cancer. *J Health Econ*. 2011;30(3):505-514. doi:http://dx.doi.org/10.1016/j.jhealeco.2011.02.001.
105. Osborne NH, Nicholas LH, Ryan AM., Humma JR, Dimick JB. Association of hospital participation in a quality reporting program with surgical outcomes and expenditures for medicare beneficiaries. *JAMA - J Am Med Assoc*. 2015;313(5):496-504. doi:http://dx.doi.org/10.1001/jama.2015.25.
106. Rajaram R, Chung JW, Jones AT, et al. Association of the 2011 ACGME resident duty hour reform with general surgery patient outcomes and with resident examination performance. *JAMA - J Am Med Assoc*. 2014;312(22):2374-2384. doi:http://dx.doi.org/10.1001/jama.2014.15277.
107. Reid RO, Ashwood JS, Friedberg MW, Weber ES, Setodji CM, Mehrotra A. Retail clinic visits and receipt of primary care. *J Gen Intern Med*. 2013;28(4):504-512. doi:http://dx.doi.org/10.1007/s11606-012-2243-x.
108. Sadhu AR, Ang AC, Ingram-Drake LA, Martinez DS, Hsueh WA, Ettner SL. Economic benefits of intensive insulin therapy in critically Ill patients: The targeted insulin therapy to improve hospital outcomes (TRIUMPH) project. *Diabetes Care*. 2008;31(8):1556-1561. doi:http://dx.doi.org/10.2337/dc07-2456.
109. Sarkar U, Lyles CR, Parker MM, et al. Use of the refill function through an online patient portal is associated with improved adherence to statins in an integrated health system. *Med Care*. 2014;52(3):194-201. doi:http://dx.doi.org/10.1097/MLR.0000000000000069.
110. Watt C, Abuya T, Warren CE, Obare F, Kanya L, Bellows B. Can reproductive health voucher programs improve quality of postnatal care? A quasi-experimental evaluation of Kenya â€™ s Safe Motherhood voucher scheme. *PLoS One*. 2015;10(4). doi:http://dx.doi.org/10.1371/journal.pone.0122828 April.
111. Domino ME, Norton EC, Morrissey JP, Thakur N. Cost shifting to jails after a change to managed mental health care. *Health Serv Res*. 2004;39(5):1379-1401. doi:http://dx.doi.org/10.1111/j.1475-6773.2004.00295.x.
112. Hodgkin D, Parks Thomas C, Simoni-Wastila L, Ritter GA, Lee S. The effect of a

three-tier formulary on antidepressant utilization and expenditures. *J Ment Heal Policy Econ*. 2008;11(2):67-77. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med5&AN=18509214; http://lib.exeter.ac.uk:4556/resserv?sid=OVID:medline&id=pmid:18509214&id=doi:&issn=1091-4358&isbn=&volume=11&issue=2&spage=67&pages=67-77&date=2008&title=The+Journal+o.

113. Li J, Hurley J, DeCicca P, Buckley G. Physician response to pay-for-performance: evidence from a natural experiment. *Health Econ*. 2014;23(8):962-978. doi:http://dx.doi.org/10.1002/hec.2971.
114. Yoon J, Bernell SL. The role of adverse physical health events on the utilization of mental health services. *Health Serv Res*. 2013;48(1):175-194. doi:http://dx.doi.org/10.1111/j.1475-6773.2012.01442.x.
117. Tannen RL, Weiner MG, Xie D. Replicated studies of two randomized trials of angiotensin-converting enzyme inhibitors: Further empiric validation of the "prior event rate ratio" to adjust for unmeasured confounding by indication. *Pharmacoepidemiol Drug Saf*. 2008;17(7):671-685. doi:http://dx.doi.org/10.1002/pds.1584.
118. Tannen R, Weiner M, Xie D. Use of Primary Care Electronic Medical Record Database in Drug Efficacy Research on Cardiovascular Outcomes: Comparison of Database and Randomised. *BMJ Br Med J*. 2009;338(b81):1-9. http://www.jstor.org/stable/10.2307/20512072. Accessed January 9, 2014.
119. Tannen D.; Wang, X.; Yu, M.; Weiner, M. G. R. X. A new "Comparative Effectiveness" assessment strategy using the THIN database: Comparison of the cardiac complications of pioglitazone and rosiglitazone. *Pharmacoepidemiol Drug Saf*. 2013;22(1):86-97. doi:http://dx.doi.org/10.1002/pds.3360.
120. Brophy S, Jones KH, Rahman MA, et al. Incidence of Campylobacter and Salmonella Infections Following First Prescription for PPI: A Cohort Study Using Routine Data. *Am J Gastroenterol*. April 2013. doi:10.1038/ajg.2013.30.
121. Uddin MJ, Groenwold RHH, van Staa TP, et al. Performance of prior event rate ratio adjustment method in pharmacoepidemiology: a simulation study. *Pharmacoepidemiol Drug Saf*. November 2014. doi:10.1002/pds.3724.
122. Lin NX, Henley WE. Prior Event Rate Ratio Adjustment for Hidden Confounding in Observational Studies of Treatment Effectiveness: A Pairwise Cox Likelihood Approach. *Stat Med*. doi:http://dx.doi.org/10.1002/sim.7051.
123. Albouy V, Lequien L. Does compulsory education lower mortality? *J Health Econ*. 2009;28(1):155-168. doi:http://dx.doi.org/10.1016/j.jhealeco.2008.09.003.
124. Swaminathan S, Mor V, Mehrotra R, Trivedi AN. Effect of medicare dialysis payment reform on use of erythropoiesis stimulating agents. *Health Serv Res*. 2015;50(3):790-808. doi:http://dx.doi.org/10.1111/1475-6773.12252.
125. Zuckerman IH, Lee E, Wutoh AK, Xue Z, Stuart B. Application of regression-discontinuity analysis in pharmaceutical health services research. *Health Serv Res*. 2006;41(2):550-563. doi:10.1111/j.1475-6773.2005.00487.x.
126. Wagner TH, Jimison HB. Computerized health information and the demand for medical care. *Value Heal*. 2003;6(1):29-39. doi:http://dx.doi.org/10.1046/j.1524-4733.2003.00155.x.
127. Fortney JC, Steffick DE, Burgess Jr JF, Maciejewski ML, Petersen LA. Are primary care services a substitute or complement for specialty and inpatient services? *Health Serv Res*. 2005;40(5):1422-1442. doi:http://dx.doi.org/10.1111/j.1475-

6773.2005.00424.x.

128. Chung S, Domino ME, Stearns SC. The effect of retirement on weight. *journals Gerontol*. 2009;Series B,(5):656-665. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed9&AN=19357073; http://lib.exeter.ac.uk:4556/resserv?sid=OVID:embase&id=pmid:19357073&id=doi:&issn=1758-5368&isbn=&volume=64&issue=5&spage=656&pages=656-665&date=2009&title=The+journa.
129. Hay J, Jhaveri M, Tangirala M, Kaliner M. Cost and resource utilization comparisons of second-generation antihistamines vs. montelukast for allergic rhinitis treatment. *Allergy Asthma Proc*. 2009;30(6):634-642. doi:http://dx.doi.org/10.2500/aap.2009.30.3293.
130. Kawatkar AA, Hay JW, Stohl W, Nichol MB. Incremental expenditure of biologic disease modifying antirheumatic treatment using instrumental variables in panel data. *Heal Econ (United Kingdom)*. 2013;22(7):807-823. doi:http://dx.doi.org/10.1002/hec.2855.
131. Piernas C, Ng SW, Mendez MA, Gordon-Larsen P, Popkin BM. A dynamic panel model of the associations of sweetened beverage purchases with dietary quality and food-purchasing patterns. *Am J Epidemiol*. 2015;181:661-671. doi:http://dx.doi.org/10.1093/aje/kwu317.