

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

36 **Abstract**

37

38 **Objective**

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

42

43 **Design**

44 Methodological review of existing literature

45

46 **Setting**

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

50

51 **Results**

52 Among the 121 studies included for review, 84 used instrumental variable analysis (IVA), of which
53 36 used lagged or historical instruments. Difference-in-differences (DiD) and fixed effects (FE)
54 models were found in 29 studies. Five of these combined IVA with DiD or FE to try to mitigate for
55 time-dependent confounding. Other less frequently used methods included prior event rate ratio
56 adjustment, regression discontinuity nested within pre-post studies, propensity score calibration,
57 perturbation analysis and negative control outcomes.

58

59 **Conclusions**

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

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

65

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

68

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

70 Word count: 199

71

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.

73 1 Introduction

74
75

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

90

91 A strength of EHRs and claims data is that they make it possible to study the comparative
92 effectiveness of interventions and the associated risk of side-effects in a real-world setting.
93 Although randomised trials provide the gold standard of evidence, observational studies based on
94 observational patient databases offer the potential to study more patients from a wider variety of
95 risk groups with a longer follow-up period at a fraction of the cost. However, in the absence of
96 randomisation, selection for treatment is often knowingly based on specific characteristics, such as
97 frailty, disease severity or the risk of an outcome. If the indication for treatment is also related to
98 prognosis, confounding by indication arises leading to biased estimation of effectiveness. There is
99 a large pharmacoepidemiologic literature on this topic and current best practice is to use design-
100 based approaches such as the Active Comparator, New User Design to help mitigate bias where
101 possible³. However, residual differences between the treatment arms other than the treatment itself
102 may still confound the intervention effect under study whether or not such an approach is used. If
103 the confounding variables are both known to the study investigators and measurable, then these
104 could potentially be adjusted for in prospective non-randomised studies. With retrospectively
105 recruited subjects, however, the recording of such variables is outside the control of the
106 investigator. Analyses of non-randomised studies that fail to account for relevant confounders may
107 have important negative consequences for health policy and patient safety.

108

109 Methods described as the quasi-experimental (QE) approach⁴, can be deployed to account for
110 confounding by unobservable characteristics. These do not attempt to directly adjust for resulting
111 bias, but use available information to achieve this indirectly under certain conditions and
112 assumptions. The aim of this systematic review is to review current practices in dealing with
113 unmeasured confounding in individual-level longitudinal health data and to capture methodological
114 developments in this area. While previous systematic reviews have been conducted to look at use
115 of propensity score methods for measured confounders^{5,6}, we are unaware of any systematic
116 review comparing use of methods for addressing unmeasured confounding in non-randomised,
117 longitudinal data. We were particularly interested in how an individual's history could be leveraged
118 to evaluate the effects of unmeasured confounding and how the extra longitudinal information
119 could be incorporated to improve adjustment for confounding bias. We intend for this review to
120 contribute to the development of best practice in addressing unmeasured confounding in
121 longitudinal data. The results should therefore help inform researchers intending to utilise “big
122 data” from electronic health records.

123

124 2 Methods

125

126 2.1 Search strategy

127

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

133

134 2.2 Inclusion and exclusion criteria

135

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

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

167 Only studies written in English were included.

168

169 The following publication types were excluded from the review:

- 170 • systematic reviews of primary studies.
- 171 • randomised controlled trials
- 172 • cross-sectional data
- 173 • preclinical and biological studies
- 174 • narrative reviews, editorials, opinions

175

176 **2.3 Study selection**

177

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

187

188 **2.4 Evidence synthesis**

189

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

195 **3 Results**

196

197 **3.1 Included studies**

198

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

202

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

206

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

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

219

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

226

227 *3.1.1 Studies excluded at full text*

228

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

245

246 **3.2 Results of the included studies**

247

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

259

260 *3.2.1 Incorporation of external/additional data*

261

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

270

271 *3.2.2 High-dimensional data*

272

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

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

282
283 *3.2.3 Quasi-experimental adjustment without longitudinal assumptions*
284

285 Those studies characterised as using a QE method without any longitudinal dimension were PSC
286 and PT as described above. We also added to this category 11 examples of Mendelian IVA^{11–21}
287 plus 32 other IVAs without historic or lagged instruments^{22–53}. While time-based instruments may
288 at first seem longitudinal, these instruments, such as date of therapy, would need to be related to
289 previous exposures or outcomes to be considered longitudinal. In some cases, survival times or rate
290 data were used, but such outcomes do not intrinsically imply longitudinal adjustment for
291 confounding. In spite of these “cross-sectional” approaches, all studies were based on some form of
292 longitudinal data at the person level, as demanded by our inclusion criteria. Among the 43 non-
293 Mendelian IVA papers in this non-longitudinal category, one study adjusted for non-longitudinal
294 fixed effects within twins³⁹. In another three, discussed below, the analysis was supplemented with
295 DiD^{38,47}, and with IVA applied to first-differences⁵⁴.

296
297 One study examined the effect of lagged, cumulative exposure to radiation on lung cancer in
298 uranium miners and nuclear workers⁵⁵. The problem of unmeasured confounding was addressed
299 using a method developed in earlier work that proposed negative control outcomes and exposures
300 as a means of both detecting and potentially resolving confounding bias⁵⁶. Here the choice of death
301 due to chronic obstructive pulmonary disorder as a negative control outcome was informed by
302 clinical knowledge of there being no direct relationship with the exposure except through the
303 possible confounder, smoking. Given a plausible negative control outcome or exposure, the method
304 offers at least a means of testing for confounding, and potentially a method of adjustment under the
305 assumption that the association between the unmeasured confounder and the negative outcome is
306 similar in magnitude to that between the same confounder and the outcome of interest.

307
308 *3.2.4 Quasi-experimental adjustment assuming time-invariant longitudinal information*
309

310 We found 36 IVA studies that used lagged information or history about the individuals’ exposure
311 as instruments^{54,57–92}. One study had recourse to the random assignment from a previous study, and
312 used this as an instrument⁶⁹. Except for that and four other different exceptions, the instruments
313 were all based at least in part on the previous intervention, or history of interventions, of the
314 clinician or healthcare facility. Characteristics of the clinician or facility may be chosen as

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

322

323 In total, 24 studies also incorporated longitudinal information through the stable of methods that, in
324 an abuse of terminology, we collectively referred to as the DiD approach. These included the 18
325 examples cited as using DiD regression^{93–110} alone, and four fixed effects (FE)^{111–114}. Either
326 through fixed effects at the individual level or through aggregate-level regression operationalizing
327 the DiD approach, these methods “ignore” the effect of confounding, which is assumed to be time-
328 invariant. At the individual level, time invariant confounding can be ignored by assigning nuisance
329 dummy variables for each individual, or cancelled out through demeaning the observations, or
330 through the first differences of observations on each individual. Two of the studies also extended
331 DiD to allow different exposure effects and trends across two-level sub-groups in the higher-order
332 contrast of difference-in-difference-in-differences^{95,106}. Fourteen studies also adjusted for
333 individual-level fixed effects either through direct inclusion of their covariates, or through
334 matching or weighting on the propensity score of the covariates. This was perhaps a more rigorous
335 and precise approach, accounting for known confounders, and yielding smaller standard errors for
336 the estimated treatment effect. However, an assumption of time-invariant confounding was still
337 required, with a null difference between exposure groups in the prior period being evidence of
338 adjustment for time-invariant confounding only. Two of the 24 DiD studies also re-analysed their
339 data using IVA^{38,47}, which provided an albeit limited opportunity to compare the relative
340 performance of these methods. In the study by Schmittdiel et al. of how statins delivered by mail
341 order affects cholesterol control⁴⁷, the intervention coefficient from modelling the single main
342 outcome was larger through DiD analysis and its standard error smaller than those from IVA, large
343 standard errors being a feature of weak instruments. The study by Lei and Lin investigated the
344 effect of exposure to a new medical scheme on 15 health outcomes and rates of health-service
345 utilisation³⁸. The effects were either not significantly different from the null or were significant and
346 of similar magnitude with similar standard error except for two outcomes, where the effect size was
347 significantly larger for IVA.

348

349 Time-invariant confounding, also known as the parallel trends assumption, was relaxed by
350 including dummy variables for the year and its interaction with the treatment dummy in a fixed-
351 effects analysis, which allowed the unobserved trend to vary between exposure groups¹¹³ using
352 methods developed in economics and therefore not captured by this review^{115,116}. The results from
353 this DiD with differential trend model were presented alongside those from the simple pooled DiD
354 model and DiD with individual fixed-effects for the effect of financial incentives in care services.
355 Tests confirmed parallel trends could be assumed in three outcomes, but out of the five outcomes
356 presented, four were statistically significant and in all, the estimated effect size by differential
357 trends was greater.

358

359 Our review also included six studies applying the prior event rate ratio method, a before-and-after
360 analogue applicable to survival and rate data¹¹⁷⁻¹²². The first two published were the seminal
361 presentation of the method applied to registry data. Also included was a comprehensive evaluation
362 by Uddin et al. of the performance of PERR under a wide array of simulated, theoretical settings,
363 under which bias was shown to increase with a greater effect of the prior events on subsequent
364 exposure or intervention. When prior events strongly influence the likelihood of treatment, the
365 exposure effect from the PERR method can be more biased than estimates from conventional
366 methods¹²¹. The problem was re-examined in a recently published study, which provided a more
367 general statistical framework for PERR adjustment and considered the potential for generalising the
368 method to allow more flexible modelling¹²².

369

370 *3.2.5 Dynamic, longitudinal quasi-experimental methods and time-varying information*

371

372 While regression discontinuity (RD) could suggest a longitudinal design, this is not exclusively so,
373 and two RD studies were excluded because of this (one applied to spatial data while the other data
374 was not longitudinal). Of those included all three could be said to accommodate time varying
375 trends¹²³⁻¹²⁵, and two of these were nested within a pre-post design: Zuckerman et al. were explicit
376 in their methodological study in identifying the robustness to time-varying confounding, in which
377 inhaler use in asthmatic patients was served as both the outcome variable in the post-test period as
378 well as the assignment variable in the pre-test period¹²⁵. In the study of the effect school-leaving
379 age on mortality by Albouy, different slopes were modelled for the assignment variable, year of
380 birth, after the cut-off date¹²³. This acknowledged different maturation rates after assignment.
381 However, as long as the assumptions of the method were met, assignment should have been as
382 good as randomised, and so no further assumptions about the temporality of confounding was
383 required.

384

385 We also picked up six examples where IVA had been combined with either DiD or a fixed effects
386 model, first appearing in our review with example from 2003¹²⁶. In Fortney's 2005 study of
387 treatment for depression¹²⁷, this combination method was justified as a control for time varying
388 confounding, referred to as second-order endogeneity. Further examples of the fixed-effects
389 instrumental variable model were found^{128,129}. The roles of lagged treatments and outcomes as
390 possible IVs and predictors were extensively considered in O'Malley's study of whether the
391 introduction of more expensive medication could have led to improved cost-effectiveness in the
392 long term⁵⁴. The author cautioned that the exclusion restriction may be difficult to satisfy when
393 using the lagged treatment as an IV after first differencing. However, two studies^{130,131} used
394 differences in the lagged explanatory variable as the IVs to adjust for second-order endogeneity in a
395 first-differences analysis following methods, not captured by our review, but developed in the
396 realm of Economics¹³²⁻¹³⁴. Referred to as the dynamic panel model or IV-GMM, this method was
397 implemented efficiently through generalised method of moments. In their report on healthcare
398 expenditure in patients with rheumatoid arthritis, Kawatkar et al. found the yielded estimates were
399 further from the null with larger standard errors when compared to those from FE alone¹³⁰.

400

401 **3.3 Implementation of methods**

402

403 While choice of method in each study often rested on which extra information was available to
404 address the issue of unmeasured confounding, method selection may also have been informed by
405 the research area. The negative control method had its origins in epidemiology, with applications to
406 occupational health policy. Likewise, the PERR method was developed exclusively on health data,
407 with applications to drug safety and public health policy. Reflecting their origins in health
408 econometrics, some studies were published in journals partially or entirely dedicated to the subject,
409 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
410 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
411 criteria, all studies had health outcomes or interventions. Mendelian IVA necessarily includes
412 genetic information, and all were published in health-related journals. In contrast, all three studies
413 using RD were published in health econometric journals.

414

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

418 either to considering the contribution from known confounders, including those summarised as a
419 propensity score, or to methods, such as perturbation testing/analysis and negative controls
420 methods, in which bias evaluation is an incremental step in adjustment. Under the assumption of
421 time-invariant confounding, the difference-in-differences method may potentially offer a way of
422 evaluating bias by modelling group differences in the pre-exposure period. However, few studies
423 evaluated hidden bias in this way^{47,96,112}. The regression formulation of the DiD method effectively
424 by-passes separate analysis of the prior period. Instead studies often discussed the within-group
425 changes over time. Similarly, the prior-period estimate from the PERR method implicitly offers an
426 evaluation of confounding bias under the same assumptions, yet none of the studies presented
427 information on outcomes in the prior period in this way. A direct evaluation of unmeasured
428 confounding is less straight-forward in IVA, with further diagnostic tests only recently developed
429 for the association between instrument and confounders^{136,137} .
430

431 **4 Discussion**

432

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

437

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

447

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

452 by choosing historical or lagged instruments. However in doing so, the causal structure needs to be
453 understood to avoid opening up “back door” paths and inducing further bias⁵⁴. DiD arguably offers
454 advantages over IVA in being more intuitive and easier to conceptualise, and with the longitudinal
455 data in EHRs it should be inherently easier to work with prior observations than to identify strong
456 instruments. Even though before-and-after methods are not subject to the imprecision of weak
457 instruments, the resulting estimates are only unbiased if the unobserved confounders exert a
458 constant effect over the observation windows before and after exposure. Where multiple
459 observations per individual exist, time may be parameterised and different trends between exposure
460 groups can be accommodated in DiD with differential trends, but a time invariant assumption about
461 confounding must still be made. To partially or wholly relax this particular assumption, instrument
462 variable analysis can be incorporated into the fixed effects model. Assuming the instrument’s
463 exclusion restriction is satisfied then this doubly-robust approach affords the advantage of DiD
464 over possibly weak instruments, while mitigating for some or all of the time-dependent
465 confounders ignored by DiD alone. Similarly, where multiple previous treatments or exposures are
466 recorded, the differenced lagged treatments can be utilised as IVs in a fixed effects model to
467 accommodate time-dependent confounding bias using the generalized method of moments system,
468 referred to as IV-GMM or the dynamic panel model.

469
470 Another potentially robust approach to unmeasured confounding would be the RD design, although
471 the small number of examples in our review probably reflects the limited number of scenarios
472 where this can be reasonably applied. Another concern over and above the usual technical
473 challenges of applying the RD method is that in spite of health records promising ample data, the
474 sample would need to be reduced to an interval around the cut-off that ensures exchangeability of
475 the two treatment groups. In this case generalisability would be restricted to individuals with
476 characteristics found in the interval. As with RD, PERR was another method that was found in
477 relatively few studies. This may have been in large part due to its recent development, rather than
478 any technically demanding aspect of its application, since it simply extends the before-and-after
479 approach of DiD to survival and rate data - outcomes that are common enough in health research.
480 However, the PERR approach does require strong assumptions including time-invariant
481 confounding and the absence of an effect of prior events on likelihood of future treatment¹²².

482
483 Methods such as IVA and DiD have their origins in the sphere of econometrics, where randomised
484 experiments are rare. We found that in importing DiD, some of the studies failed to explicitly
485 acknowledge the problem of confounding bias. Instead justification for the method was presented
486 in terms of the common trends assumption. Discussion of possible confounding bias is regarded as

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

494

495 Choosing between methods to reduce unmeasured confounding bias is challenging and we found
496 few studies that directly compare methods. The performance of different methods will depend on
497 factors such as the nature of the underlying confounding, the type of exposure and outcome, and
498 the sample size¹³⁸. The type of data available will also guide the choice of method. For example,
499 the instrumental variable method requires a suitable instrument and DiD / PERR require data on at
500 least two periods. In practice, no one method is likely to be best suited to all problems, and it is
501 essential for investigators to carefully assess the potential biases in each proposed study, where
502 possible tailoring the methods or combination of methods to address these biases¹³⁹. Our review
503 has highlighted how use of longitudinal information is one additional and potentially important
504 consideration in this process.

505

506 While our review focussed on the problem of adjustment using analytic methods, many problems
507 associated with observational data may be pre-empted by use of an appropriate study design¹⁴⁰.
508 Before choosing an appropriate analytic method, it is recommended that investigators carefully
509 identify and match individuals for the control and intervention groups in order not to exacerbate
510 any bias³. The importance of study design is often discussed with a view to minimising
511 confounding bias from unmeasured sources, with the subsequent adjustment accounting for
512 observed confounders only¹⁴¹, usually through the matching, weighting or adjustment of propensity
513 scores¹⁴². Where the success of the design remains in doubt, or its criteria cannot be fully met, then
514 investigators will inevitably need recourse to some of the alternative methods reviewed in this
515 report.

516

517 The reviewed studies did not seek to distinguish between the different mechanisms of bias.
518 Confounding by indication, deemed intractable by many researchers using the observed data¹⁴³,
519 was seen to create additional sources of bias in two separate simulation studies applying the
520 “longitudinal” method of PERR, when an association was modelled between prior events and
521 treatment status in the study period^{121,122}. Another common form of selection bias in

522 pharmacoepidemiologic studies is the healthy user bias and this works in the opposite direction to
523 confounding by indication, distorting treatment-outcome associations towards the treatment
524 looking beneficial³. Further research is needed to understand how each of the methods in this
525 review is affected by the different types of confounding.

526

527 An inherent limitation of this large, wide-ranging review is that it precluded meaningful data
528 synthesis due to the mix of different data and study types. Furthermore, we could only find a few
529 examples where the performance of different methods was compared within the same study. We
530 also stipulated in the inclusion criteria that unmeasured confounding, or any of its synonyms,
531 should be given as justification for methods in its adjustment. This may have inadvertently
532 excluded some papers, where justification was implicit, but good practice in health research
533 demands acknowledgement of this source of bias where applicable. While our search terms were
534 specific to the scope of our review, we accept that this may have inadvertently excluded relevant
535 methods and studies. Some methods, such as negative control outcomes, that were identified in the
536 original search were not included as explicit terms in the search strategy, and further secondary
537 searches may have uncovered additional studies using these methods. We also acknowledge that
538 there may be other relevant methods for addressing unmeasured confounding that have been missed
539 by the search strategy. Consequently, we made inferences about the relative application of methods
540 with caution. However, we were surprised so many studies focussed solely on IVA as the sole
541 means of adjustment. A similar conclusion was echoed by a different review on regression
542 discontinuity designs that found interest was growing in RD only as recently as 2014¹⁴⁴.

543

544 By choosing to focus on methods with an independent control arm for each treatment, our review
545 excluded case only designs including case-crossover designs (CCO) and the self-controlled case-
546 series design. This class of methods addresses unmeasured confounding by making comparisons
547 within individuals so that each individual acts as his or her own control. Another case-only design,
548 the case-time control design, is an extension of the CCO design that uses information from a
549 historical control group in a similar way to the PERR method. These approaches are reviewed by
550 Uddin et al¹³⁸ and Nordmann et al¹⁴⁵.

551

552 This review has considered a range of promising new methods for addressing unmeasured
553 confounding in non-randomised studies. However, consistent with prior research on dissemination
554 and uptake of statistical innovations¹⁴⁶, the rate of knowledge translation has been slow and we
555 found that most studies in our review used established methods such as IVA and DiD. A recent
556 study by Cadarette et al has shown how Rogers' Diffusion of Innovations model can be used to

557 describe the adoption of novel methodologies in pharmacoepidemiology¹⁴⁷ and this provides a
558 useful resource for interpreting the uptake of methods in this review. Cadarette et al proposed five
559 principles for authors of methodological innovations that may improve translation into practice¹⁴⁷:
560 (1) clearly describing the methods using foundational principles; (2) comparing results to
561 established methods; (3) providing sample data, code or calculation examples; (4) early
562 communication, support and testing; and (5) providing methodological and reporting guidance.
563 These recommendations offer a useful checklist for researchers developing methods for addressing
564 unmeasured confounding in observational studies. Of particular relevance in the context of this
565 review is the need for more extensive evaluation and comparison of the emerging methods in a
566 range of settings. The review also addresses the need for methodological guidance through
567 highlighting the potentially important role of longitudinal information in addressing confounding
568 bias and has identified this as an area for further development.

569 **5 Conclusions**

570
571 Our review showed how seminal work in econometrics has influenced practice in dealing with
572 unmeasured confounding in clinical and epidemiological research. Although the issue of
573 unmeasured confounding is widely acknowledged, we found that longitudinal information in
574 observational studies appears under-utilised. Lagged and historical characteristics associated with
575 the treatment may help enforce the exclusion restrictions of instrumental variables under the
576 appropriate causal structures, while before-and-after methods, such as DiD and PERR, afford an
577 intuitive approach without the imprecision of weak instruments. Furthermore, they offer a direct
578 evaluation of time-invariant confounding bias. The most robust methods we found applied
579 instrumental variable analysis to the fixed effects difference-in-differences method, where such
580 suitable instruments or difference lagged variables could be assumed to satisfy the exclusion
581 restriction. While there are sometimes good technical reasons for choosing one mode of analysis
582 over another, many questions remain over the most appropriate methods. All methods rely on
583 assumptions, but little guidance is available to applied researchers as to the empirical settings in
584 which particular methods can be safely used. Few studies directly compare different methods and
585 more research is needed to establish the relative performance of the methods in realistic
586 settings.

587

588 Acknowledgements

589 This research was funded by the Medical Research Council [grant number G0902158] with
590 additional support from the National Institute for Health Research (NIHR) Collaboration for
591 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.

602 **References**

603

- 604 1. Murdoch TB, Detsky AS. The inevitable application of big data to health care. *JAMA*.
605 2013;309(13):1351-1352. doi:10.1001/jama.2013.393.
- 606 2. Safran C, Bloomrosen M, Hammond WE, et al. Toward a national framework for the
607 secondary use of health data: an American Medical Informatics Association White
608 Paper. *J Am Med Inform Assoc*. 2007;14(1):1-9. doi:10.1197/jamia.M2273.
- 609 3. Lund JL, Richardson DB, Stürmer T. The Active Comparator, New User Study Design in
610 Pharmacoepidemiology: Historical Foundations and Contemporary Application. *Curr*
611 *Epidemiol Reports*. 2015;2(4):221-228. doi:10.1007/s40471-015-0053-5.
- 612 4. Cook TD, Campbell DT. *Quasi-Experimentation: Design & Analysis Issues for Field*
613 *Settings*. 3rd ed. Chicago: Rand McNally, 1979; 1979.
614 <http://lib.exeter.ac.uk/record=b1090074~S6>. Accessed January 13, 2015.
- 615 5. Shah BR, Laupacis A, Hux JE, Austin PC. Propensity score methods gave similar results
616 to traditional regression modeling in observational studies: a systematic review. *J Clin*
617 *Epidemiol*. 2005;58(6):550-559. doi:10.1016/j.jclinepi.2004.10.016.
- 618 6. Stürmer T, Joshi M, Glynn RJ, Avorn J, Rothman KJ, Schneeweiss S. A review of the
619 application of propensity score methods yielded increasing use, advantages in specific
620 settings, but not substantially different estimates compared with conventional
621 multivariable methods. *J Clin Epidemiol*. 2006;59(5):437-447.
622 doi:10.1016/j.jclinepi.2005.07.004.
- 623 56. Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative Controls. *Epidemiology*.
624 2010;21(3):383-388. doi:10.1097/EDE.0b013e3181d61eeb.
- 625 115. Wagstaff A, Moreno-Serra R. Europe and central Asia's great post-communist social
626 health insurance experiment: Aggregate impacts on health sector outcomes. *J Health*
627 *Econ*. 2009;28(2):322-340. doi:10.1016/j.jhealeco.2008.10.011.
- 628 116. Bell B, Blundell R, Reenen J Van. Getting the Unemployed Back to Work: The Role of
629 Targeted Wage Subsidies. *Int Tax Public Financ*. 1999;6(3):339-360.
630 doi:10.1023/A:1008787013977.
- 631 132. Arellano M, Bond S. Some Tests of Specification for Panel Data: Monte Carlo Evidence
632 and an Application to Employment Equations. *Rev Econ Stud*. 1991;58(2):277.
633 doi:10.2307/2297968.
- 634 133. Arellano M, Bover O. Another look at the instrumental variable estimation of error-
635 components models. *J Econom*. 1995;68(1):29-51. doi:10.1016/0304-4076(94)01642-
636 D.
- 637 134. Blundell R, Bond S. Initial conditions and moment restrictions in dynamic panel data
638 models. *J Econom*. 1998;87(1):115-143. doi:10.1016/S0304-4076(98)00009-8.
- 639 135. MacKenzie TA, Tosteson TD, Morden NE, Stukel TA, O'Malley AJ. Using instrumental
640 variables to estimate a Cox's proportional hazards regression subject to additive
641 confounding. *Heal Serv Outcomes Res Methodol*. 2014;14(1-2):54-68.
642 doi:10.1007/s10742-014-0117-x.
- 643 136. Jackson JW, Swanson SA. Toward a clearer portrayal of confounding bias in
644 instrumental variable applications. *Epidemiology*. 2015;26(4):498-504.
645 doi:10.1097/EDE.0000000000000287.
- 646 137. Davies NM. An even clearer portrait of bias in observational studies? *Epidemiology*.
647 2015;26(4):505-508. doi:10.1097/EDE.0000000000000302.
- 648 138. Uddin MJ, Groenwold RHH, Ali MS, et al. Methods to control for unmeasured
649 confounding in pharmacoepidemiology: an overview. *Int J Clin Pharm*. April 2016.
650 doi:10.1007/s11096-016-0299-0.
- 651 139. Alemayehu D, Alvir JM, Jones B, Willke RJ. Statistical issues with the analysis of
652 nonrandomized studies in comparative effectiveness research. *J Manag Care Pharm*.

653 2011;17(9 Suppl A):S22-6. <http://www.ncbi.nlm.nih.gov/pubmed/22074671>.
654 Accessed July 18, 2014.

655 140. Velentgas P, Dreyer NA, Nourjah P, Smith SR, Torchia MM. *Developing a Protocol for*
656 *Observational Comparative Effectiveness Research: A User's Guide*. Agency for Healthcare
657 Research and Quality (US); 2013. <http://www.ncbi.nlm.nih.gov/pubmed/23469377>.
658 Accessed October 25, 2016.

659 141. Freemantle N, Marston L, Walters K, Wood J, Reynolds MR, Petersen I. Making
660 inferences on treatment effects from real world data: propensity scores, confounding
661 by indication, and other perils for the unwary in observational research. *Bmj*.
662 2013;347(nov11 3):f6409-f6409. doi:10.1136/bmj.f6409.

663 142. Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of
664 Confounding in Observational Studies. *Multivariate Behav Res*. 2011;46(3):399-424.
665 doi:10.1080/00273171.2011.568786.

666 143. Bosco JL, Silliman RA, Thwin SS, et al. A most stubborn bias: no adjustment method
667 fully resolves confounding by indication in observational studies. *J Clin Epidemiol*.
668 2010;63(1):64-74. doi:<http://dx.doi.org/10.1016/j.jclinepi.2009.03.001>.

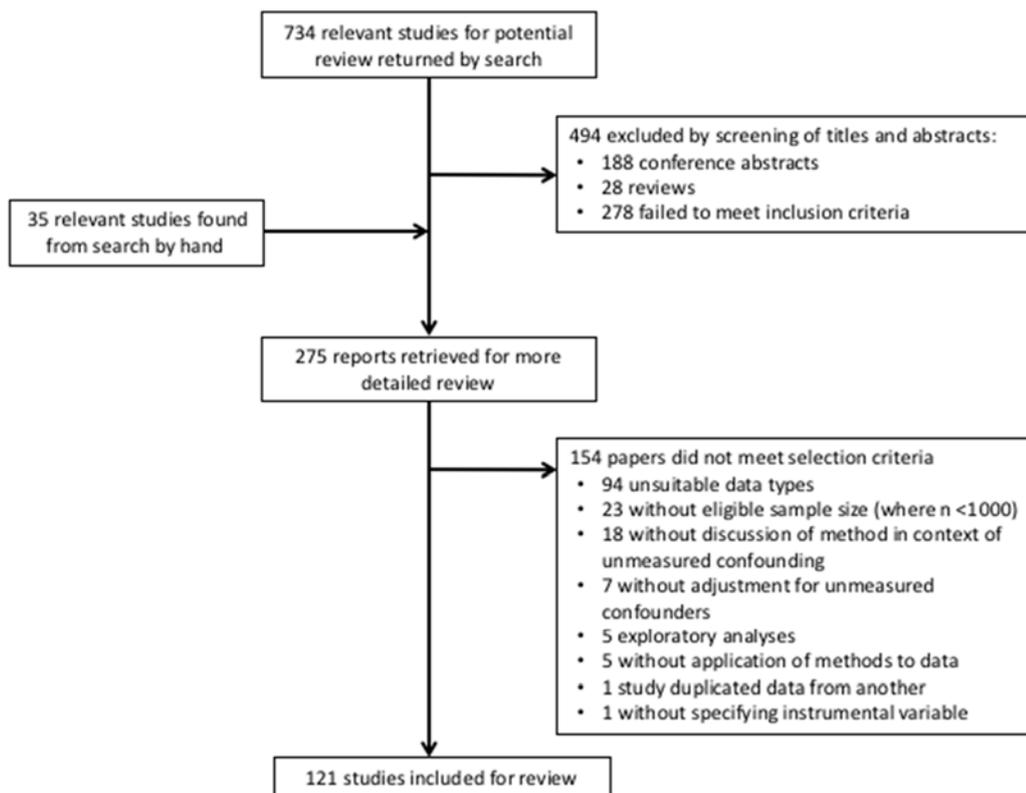
669 144. Moscoe E, Bor J, Bärnighausen T. Regression discontinuity designs are underutilized in
670 medicine, epidemiology, and public health: a review of current and best practice. *J Clin*
671 *Epidemiol*. 2015;68(2):122-133. doi:10.1016/j.jclinepi.2014.06.021.

672 145. Nordmann S, Biard L, Ravaud P, et al. Case-Only Designs in Pharmacoepidemiology: A
673 Systematic Review. Little J, ed. *PLoS One*. 2012;7(11):e49444.
674 doi:10.1371/journal.pone.0049444.

675 146. Pullenayegum EM, Platt RW, Barwick M, Feldman BM, Offringa M, Thabane L.
676 Knowledge translation in biostatistics: a survey of current practices, preferences, and
677 barriers to the dissemination and uptake of new statistical methods. *Stat Med*.
678 2016;35:805-18.

679 147. Cadarette SM, Ban JK, Consiglio GP, Black CD, Dubins D, Marin A, Tadrous M. Diffusion
680 of Innovations model helps interpret the comparative uptake of two methodological
681 innovations: co-authorship network analysis and recommendations for the integration
682 of novel methods in practice. *J Clin Epidemiol*. 2016 Dec 23. pii: S0895-4356(16)30835-
683 6.
684
685

686 **Figures**
687



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

690

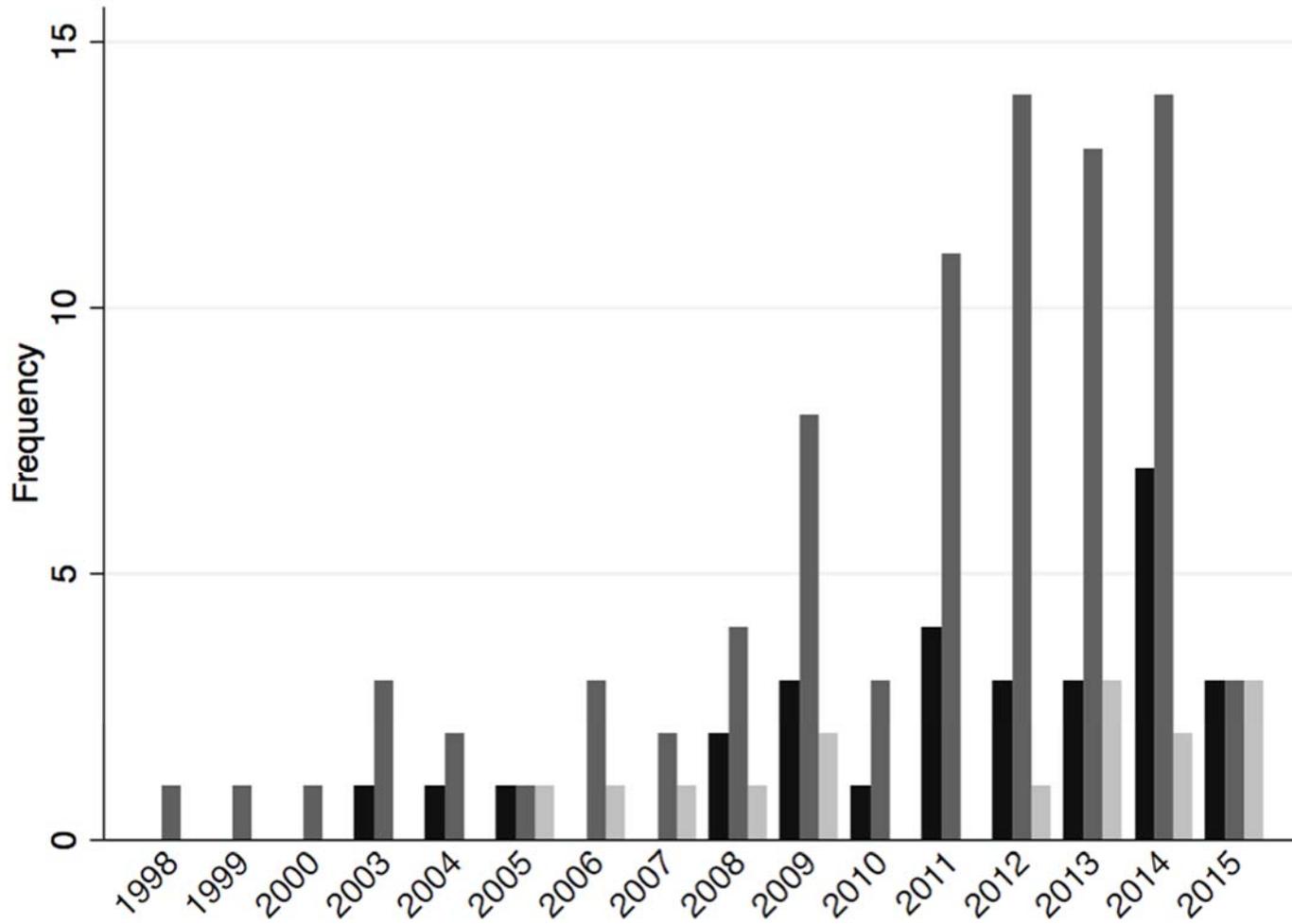


Figure 2: Plot of frequency of reviewed methods for mitigating for unmeasured confounding by: difference-in-differences [black]; Instrumental variable analysis (IVA) [mid-grey]; Other [light grey] includes regression discontinuity, prior event rate ratio method, propensity score calibration, perturbation 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 control is 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	1	20
Time	Time-based characteristic of treatment such as date of therapy	6			10
Historical	Usually prescribing preference of physician or facility based on historical records of previously administered therapies	31	2	1	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	
Goetzl, 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.; Lavalley, 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.; Crespino, 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.; Dorransoro, 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