Published in final edited form as:
J Clin Epidemiol. 2011 June ; 64(6): 687–700. doi:10.1016/j.jclinepi.2010.09.006.

Use of Instrumental Variable in Prescription Drug Research with Observational Data: A Systematic Review

Yong Chen[a,b] and **Becky A. Briesacher**[a]

[a]Meyers Primary Care Institute and University of Massachusetts Medical School, Worcester, MA

[b]Clinical and Population Health Research Doctoral Program, Graduate School of Biomedical Sciences, University of Massachusetts Medical School, Worcester, MA

Abstract

Objective—Instrumental variable (IV) analysis may offer a useful approach to the problem of unmeasured confounding in prescription drug research if the IV is: 1) strongly and unbiasedly associated to treatment assignment; and 2) uncorrelated with factors predicting the outcome (key assumptions).

Study Design and Methods—We conducted a systematic review of the use of IV methods in prescription drug research to identify the major types of IVs and the evidence for meeting IV assumptions. We searched MEDLINE, OVID, PsychoInfo, Econlit and economic databases from 1961 to 2009.

Results—We identified 26 studies. Most (n=16) were published after 2007. We identified five types of IVs: regional variation (n=8), facility prescribing patterns (n=5), physician preference (n=8), patient history/financial status (n=3) and calendar time (n=4). Evidence supporting the validity of IV was inconsistent. All studies addressed the first IV assumption; however, there was no standard for demonstrating that the IV sufficiently predicted treatment assignment. For the second assumption, 23 studies provided explicit argument that IV was uncorrelated with the outcome, and 16 supported argument with empirical evidence.

Conclusions—Use of IV methods is increasing in prescription drug research. However, we did not find evidence of a dominant IV. Future research should develop standards for reporting the validity and strength of IV according to key assumptions.

1. Introduction

Evidence-based medicine is essential to assure that effective and safe medications are prescribed for the right reasons to the right individuals. In the best case, prescribing decisions are based on current medical evidence. However, an Institute of Medicine (IOM) report indicates that more than half of all treatment provided in the United States is not supported by evidence.[1] This is especially true for vulnerable patient populations who are under-represented in randomized clinical trials, such as the elderly and the frail [2,3], and for comparisons between therapeutic alternatives rather than between active therapy and placebo.

Address: Yong Chen, CPHR Graduate Student, University of Massachusetts Medical School, Biotech Four, Ste 315, 630 Plantation St, Worcester, MA 01605, United State, Phone: +1 508 856 2611, Fax: +1 508 856 5024, yong.chen@umassmed.edu.

The obvious solution of filling these gaps with evidence from observational research is greatly diminished, however, by the strong effects of confounding by indication. Individuals who receive therapies are different from those who do not, and a simple comparison of the treatment effects without accounting for these differences will produce biased results. Instrumental variable (IV) approach is a potential method for addressing measured and unmeasured confounding in observational studies [4].

1.1 Significance of IV analysis in prescription drug research

The IV method has been used in the social sciences and economics fields for decades but it was introduced to the medical literature in 1989 [5]. To date, IV methods have been applied to a wide range of medical intervention research questions [6-12].

IV analyses may be particularly relevant for prescription drug research that uses large, secondary data sources to compare the effectiveness of two medications, or to examine the effects of medications in special patient populations. The IV approach may be a preferred approach if the unmeasured confounding is expected to be significant and the IV is strong and valid.

1.2 Objectives

While IV analyses are emerging in the medical research field, the extent to which this technique has been adopted in prescription drug research is not known. The objective of this paper is to systematically review the medical literature on the use of IV analysis in prescription drug research. Specifically, we identified: 1) the frequency of research using IV analysis over time; 2) a list of candidate IVs and 3) the evidence for the validity of the candidate IVs.

2. Instrumental variable

2.1 Background

IV analysis is a technique enabling researchers to take advantage of observational data such as claims data and registry data to more correctly estimate the effectiveness or safety of a medication even if unmeasured risk factors are present. Figure 1, adopted from Brookhart et al., [13] illustrates this technique. The central idea of IV analysis is to find a variable that is strongly associated with the treatment assignment, in this case a prescription drug, but is not related to the outcome, except through its relationship with the treatment assignment.[4,14] A good IV should satisfy two key assumptions. First, the IV should be strongly related to the treatment assignment and this association should be estimated without bias. Second, the instrument should not be correlated with measured and unmeasured confounders and only related to the outcome through the treatment assignment (exclusion criteria). [14] This means that the IV should neither be related to risk factors of the outcome (the uppermost pathway [dash line] in Figure 1) nor have direct effect on the outcome (the lowermost pathway [dash-dot line] in Figure 1). Therefore, it is related with the outcome only through the treatment assignment (the middle pathway [dot line] in Figure 1).

2.2 Examples of IV and IV estimator

A coin toss randomizing patients to the treatment or placebo arms in a clinical trial is an example of a perfect IV. Treatment assignment is perfectly correlated with the disposition of the coin toss, so it meets the first IV assumption. Also, the coin toss is completely independent of the outcome, except through treatment assignment. This qualifies the coin toss for the second IV assumption.

In another example, Brookhart and colleagues[15] calculated a measure of physician preference for Cyclooxygenase 2 inhibitor (COX-2 inhibitor) over non-selective non-steroidal anti-inflammatory drugs (NSAID) as the IV. In that study of COX-2 inhibitor use and risk of GI complications, the investigators argued that patients are more likely to receive COX-2 inhibitors if their physician prefers these agents. Furthermore, this prescribing preference supersedes the patient's indication for this medication, and patients will not select physicians based on the prescribing preference. These arguments suggest that physician's treatment preference is a valid IV.

The IV approach is summarized in this equation

$$\frac{E[Y|Z=1] - E[Y|Z=0]}{\Pr[X|Z=1] - \Pr[X|Z=0]} \qquad \text{(Equation 1)}$$

where X and Y represent the treatment and outcome respectively and Z represents the IV. E[Y|Z] is the expected value of outcome given Z, the IV, while Pr[X|Z] represents the probability of exploratory treatment variable given Z.

Using the previous example, Y is equal to 1 if the outcome, a GI complication, is present and 0 otherwise. X represents treatment assignment with 1 indicating receipt of a COX-2 inhibitor and 0 otherwise. Z is the physician's preference with 1 indicating a preference for COX-2 inhibitor and 0 otherwise. E [Y|Z=1] represents the expected value of outcome given that the physician prefers COX-2 inhibitors. Pr[X|Z=1] is equal to the probability of receiving a COX-2 inhibitor given that a physician's preference is COX-2 inhibitors. If the IV is perfect, 100% of patients whose physician prefers COX-2 inhibitors will receive a COX-2 inhibitor and 0% of patients whose physician prefers NSAIDs will receive a COX-2 inhibitor. The denominator of IV estimator (equation 1) reduces to 1, and comparative effectiveness is merely the difference in outcome between patients receiving COX-2 inhibitors and patients receiving NSAIDs. Theoretically, the IV estimate is equal to the results from a RCT.

3. Methods

3.1 Search strategy

We searched MEDLINE, OVID, PsychoInfo, Econlit and National Bureau for Economic Research (NBER) databases from 1961 to 2009 using key terms, (*prescription drugs* OR *medication* OR *treatment* OR *medicine*) AND (*instrumental variable*); 785 articles were identified. We reviewed titles and abstracts obtained from the search and excluded studies that did not meet the inclusion and exclusion criteria. Finally, we reviewed the bibliographies of the included articles to identify additional articles.

3.2 Inclusion and exclusion criteria

We included a study if: 1) it was published in an English language peer-reviewed journal; 2) prescription drug was used as the exploratory treatment variable (exposure); 3) a patient outcome that is related to the treatment was investigated; and 4) IV analysis was the main analytical approach, meaning that the authors described the IV method and reported the results.

We excluded a study if: 1) it emphasized on pure methodological or statistical research with simulated data; 2) it was a review, abstract, book chapter or dissertation; and 3) IV analysis was used in a clinical trial.

3.3 Data abstraction

We developed a data abstraction form followed by a pilot test of the form. Data were recorded on abstract forms and stored in excel files. First, we captured information such as a description of the study year, design, population, exploratory treatment variable, adjusted confounders and outcome. Second we abstracted detailed data on IV including type and number of IVs, and analytical approaches. Finally, we reviewed whether and how the two key assumptions were verified in the studies.

3.4 Quality rating

We did not assess the overall quality of the selected studies because none of the popularly used approaches, such as Downs and Black, QUOROM and MOOSE, is designed for assessing a methodology. However, we did assess the quality of each candidate IV based on the two criteria of a valid IV. We assigned a quality score of 2 if the paper discussed or provided empirical evidence for the two key assumptions; 1 if only one of the two assumptions was discussed, and 0 if none of the assumptions were discussed. Two investigators developed the search strategy (Y.C, and B.B.), 1 (Y.C.) retrieved the articles and extracted the data, 2 (Y.C., B.B.) made the final selection of studies. One author (Y.C.) assigned the initial quality scores, and two authors (Y.C., B.B.) reached consensus on the final score.

3.5 Analysis

We computed the number of studies containing IV analysis for prescription drug research across time and presented the distribution of IV studies across different 3-year periods (2001-2003, 2004-2006 and 2007-2009) to identify whether there was a trend of IV analysis in prescription drug research. We grouped studies by types of IVs and the key assumptions, and computed the number or percentage of studies in different groups.

4. Results

Our search identified 785 papers (Figure 2). We initially excluded 486 articles as irrelevant or published in non-English language journals. In the remaining 299 studies, we excluded duplicates (n=97), reviews or summaries (n=17), dissertations (n=9), book chapters (n=3) and abstracts (n=3). We also excluded studies that used prescription drug as the outcome but not the exploratory treatment variable (n=3); adopted RCT designs (n=4); were pure methodological or statistical studies with simulated data (n=27) or the exploratory variable was not prescription drug (n=111). Thus, the remaining 25 possible articles were reviewed in detailed and none of them was excluded after review. One additional article was identified from a manual review of the bibliographies for a total of 26 articles.

The first study that used IV analysis in prescription drug research was published in 2001[8] (Table 1). Since then, the rate of publications has been increasing over time, from two studies (2001-2003) to eight (2004 to 2006) to 16 studies (2007-2009).

Table 1 shows the characteristics of the studies, which were conducted tin Japan, Canada., and the United States. The sample size ranged from 336 to 170,024. Most of the studies were conducted in older age populations: 15 were conducted in patients older than 55 years; [8,15-28] and another two studies described their populations having mean age of greater than 60 years [29,30]. Only one study was conducted in a pediatric population.[31]

Table 2 describes detailed information on the IVs, the exploratory treatment variable and the outcome for all studies. The IVs generally fell into five categories: regional variation, facility prescribing patterns, physician preference, patient history/financial status, calendar

time and others. Regional variation IV was adopted in 8 studies.[8,19,28,31-35] Most studies used the proportion of patients on treatment (e.g. antihypertensive medication[33]) in a region (e.g. health care service area [HCSA][19]) as an IV. Five studies considered facility-prescribing patterns as an IV [17,20,30,34,35]. Similar to regional variation, facility-prescribing patterns was measured as the proportion of patients on a treatment within a facility. Physician preference was another popularly used IV with 8 studies deriving IV from physician preference.[15,16,18,21-23,26,36] It can be measured by either a physician's last prescription or the proportion of patients on treatment. The IV derived from patient history/financial status was included in 3 studies.[18,25,27] For instances, one study used patient medication coverage and the others used patient medical history (e.g. history of gout [18,27]) as IVs. Calendar time was used as an IV in 4 studies [24,37-39] (e.g. before and after 1996 when highly active antiretroviral therapy [HAART] was available [39]). Only one study used propensity score as an IV,[29] although the usage of propensity score as an IV may be questionable because the study propensity score only controlled for measured confounders. While an individual IV was found in 23 studies, three studies [32,34,35] adopted multiple IVs in the analysis.

Studies with IV analyses covered a wide range of prescription drugs. Most studies involved a specific group of medication (Table 2) (i.e. antipsychotic, [21,22,26,34,35]; NSAID, [15,23] and HAART [32,39]); others investigated combinations of drugs[17] or prescription drugs in general [25]. Comparisons were made between use prescription drugs nondrug alternatives [18,27,31-33,39] and between medications and active comparators. [15,20,21,23,34-36]

Two types of outcomes were generally studied with IV analyses, treatment effectiveness and adverse drug event. Effectiveness research was performed in 17 studies [8,16-19,25,27-29,31-35,37-39]. For instance, Cain et al.[39] examined the effectiveness of HAART on the development of clinical acquired immune deficiency syndrome (AIDS) in human immunodeficiency virus positive (HIV+) patients. Nine studies evaluated adverse drug events [15,20-24,26,30,36]. For instance, in a study conducted by Brookhart et al.[15] selective COX-2 inhibitors were compared with NSAIDs in term of GI complications among patients older than 65.

Table 3 presents the analytical approaches of IV analysis, and how the two key assumptions were addressed. More than half of the studies (n=15) adopted a two stage model such as two stage least square (2SLS), [15,21-23,26,31-33,36] two stage residual inclusion (2SRI) [25,38]. A Probit structured equations model was used in 3 studies.[27,32,35] One study constructed a three stage least square (3SLS) [24]. Since the IV estimator includes two parts (Equation 1), some researchers estimated both parts separately and combined them into a single IV estimator and confidence interval was then estimated by the bootstrapping methods in 5 studies [8,18,19,28,39]. A propensity score was used as IV and adjusted in one study [29]. The analytical approach was not abstracted for one study due to the unclear description of the study method [17].

There was a considerable heterogeneity in discussions of the two key assumptions of IV. All studies (Table 3) explained why the selected IV was associated with the exploratory treatment variable (assumption 1). Some authors modeled the exploratory treatment variable as a function of the IV and reported the regression coefficient or odds ratio and p value. For instance, Bosco et al.[16] performed linear regression to test the strength of association between IV (physician's preference of adjuvant chemotherapy) and treatment assignment (adjuvant chemotherapy) and reported that the coefficient was 23.7%. Some authors presented F-statistic or R-square from the two-stage model and probit structure equation. For instance, Zhang[38] compared the effect of two bipolar disorder medications using a 2SRI

and reported that the first stage F-statistic was equal to 27.63 ($P<0.05$) indicating a strong instrumental variable. Others provided data indicating that the probability of receiving treatment varied among different categories of IV. For instance, Yao-Lu et al.[19] reported that the probability of receiving primary androgen deprivation therapy for a patient in HSAs with the highest prevalence of PADT was 53%, while the probability was 31% for a patient in the healthcare service area with the lowest prevalence of PADT. [26] We found that less than half of the studies (n=11) [8,15,16,19,20,23,26,28,31,32,38] explicitly reported whether the IV induced variation.

The second assumption that IV is not independently correlated with the outcome was discussed in 23 of the 26 studies. This assumption is unverifiable; however, some researchers provided exploratory evidence. Two studies [12,26], reported that their IV was associated with the measured confounders. But most of the studies (n= 13) [8,15,16,19-21,23,26,28,30,31] reported that IV analysis attenuated the imbalance of the observed confounders between treatment and non-treatment groups. Some authors argued that the IV was valid if it was able to reduce the imbalance of observed confounders.[13,31] Furthermore, some authors examined whether the IV had a direct effect on the outcome [34,35]. A number of studies verbally argued why their IVs would not be related to measured and unmeasured confounders. For instance, Schneeweiss et al. used a surgeon's last antifibrinolytic agent as the IV and argued that patients were unlikely to choose their surgeon on the basis of the surgeon's preference for a specific agent.[36] Lastly, 3 studies [22,27,37] did not discuss the second IV assumption.

Table 3 describes the quality of IVs based on satisfaction of the two key assumptions. Overall, 17 studies received a quality score of 2 while a score of 1 was assigned to 9 studies. No study received a score of 0 since all studies assessed and provided rationale for the first assumption. The reasons that an IV received a quality score of 1 were that 1) IV was associated with measured confounders (n=6)[15-18,29,39] and 2) the authors did not discuss the second assumption (n=3).[22,27,37]

Since the IV method is designed to mimic the RCT, it is worthwhile to note when results from IV analysis were comparable to the RCT results. Sixteen studies [8,15,16,19-21,23-25,27,28,30,33-36] reported that RCTs were available at the time of their studies. Fifteen studies found that results from IV analysis were consistent with those from RCTs, although one study[16] found that the IV analysis and RCT were inconsistent due to a weak IV. Furthermore, 13[8,15,19-21,23,25,27,28,30,34-36] out of these 16 studies also reported using other methods such as survival analysis, linear regression etc. beside IV analysis. Nine of them[15,19,23,25,27,28,30,34,35] found that in contrast to IV analysis, other methods provided estimates inconsistent with those from RCT suggesting IV yielded better control of measured and unmeasured confounders.

5. Discussion

The use of IV analyses in the medical literature has been growing since the landmark publication by McClellan et al. [11] in 1994. However, the use of IV methodology remains limited and relatively new in prescription drug research. Our systematic review identified only 26 studies, with the earliest from 2001 and the majority occurring after 2006. Possible reasons for this slow adoption of this method include unfamiliarity with IV analyses and the practical difficulty of finding valid IVs.[5].

However, IV analyses have the potential to fill important gaps in evidence-based medicine. Compared with the RCT, observational studies using IV analysis offer expansions in generalizability to under-represented and small sample populations [2,3]. We found that IV

analyses were common in older age populations, which have historically been excluded from many RCTs. The sample size in IV studies was also generally larger than the typical RCT. For instance the Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE), one of the largest post market trials for comparing the effectiveness of antipsychotics, recruited only 1,494[40] and 421[41] participants with schizophrenia and dementia respectively. In comparison, Wang et al.[26] used IV analyses in studying the effectiveness of atypical vs. typical antipsychotics among 22,890 patients. However, we only found one IV study in a pediatric population, a very difficult population in which to conduct RCTs. This suggests that IV analysis for prescription drug research may be expanded to other vulnerable populations. Furthermore, in contrast to RCTs, observational studies frequently compare a medication with its active comparators as more than 40% of the studies (n=11) in this review involved at least one active comparator.

We observed that IVs for prescription drug research generally fall into 5 categories: regional variation, facility-prescribing patterns, physician preference, patient history/financial status and calendar time. The first three categories were especially common in this review. A possible reason may be the ease of computing these variables using administrative data. Furthermore, the IV could be tailored to address a wide variety of clinical issues. In should be noted though that research has found that these computed IVs are specific to the source population. For instance, Brookhart et al.[15] and Schneeweiss [21] use physician's prescribing preference as an IV for a U.S. and Canada study. They found evidence that prescribing preference might be sensitive to particular health care systems and geographic regions.

We also found heterogeneity in this review of the discussion and verification of the key IV assumptions. All studies assessed the first assumption that IV, however, some did not address the strength of the IV. An IV that weakly predicts treatment assignment may exaggerate bias. [4,42-44] This occurs when small IV-induced variation leads in denominator of the IV estimator (equation 1) magnifies the unmeasured confounding in the numerator [42]. Furthermore, the size of the marginal population depends on the magnitude of the IV induced variation [4]. We found that less than half of the studies (n=11) [8,15,16,19,20,23,26,28,31,32,38] explicitly reported IV induced variation. Furthermore there was no consensus regarding the magnitude of the induced variation that may leads to bias. For instance, Brookhart et al.[15] argued that their IV was strong because induced variance in their study was 23% (53%-30%). However, Hernan et al.[42] argued that this was a weak IV because the IV estimator would be exaggerated by 4.4 times (1/0.23). Therefore, we suggest that future research explicitly report both strength of the association between IV and exploratory treatment variable, and IV induced variation in order for readers to justify the strength of the IV.

There was a substantial heterogeneity in terms of the assessment of the second assumption, and three studies did not mention it. Only 23 studies provided an explicit argument that IV was uncorrelated with the treatment selection with reasons. Among these 23 studies, 16 supported the argument with empirical evidence such as reporting data that IV was not directly related to the outcome (n=3) and reporting that the IV analysis reduced the imbalance of measured patient-level factors between the treatment and non-treatment groups implying that the IV also reduced the imbalance of unmeasured patient factors and other empirical evidence (n=13). Therefore more a detailed assessment of the second assumption is needed to assure the validity of IV analysis in future studies.

This review is subject to limitations. First, we focused our review on prescription drug research. The discussion of IV analysis may not be generalized to other medical treatment. Second, the increasing trend of using IV analysis for prescription drugs may be due to

increasing familiarity of this method to clinical research or to an increase of prescription drugs on the market. Third, there is not a consensus regarding the appropriateness of different analytical approaches for IV. Therefore, we did not assess the statistical methods in this review. Fourth, we reported the consistency of results between RCT and IV analysis, although they might not be directly comparable as the estimates are drawn from different populations.

Nevertheless, our empirical assessment of the literature demonstrates that researchers may identify a valid IV, certainly from among the five types of IVs summarized in this paper. However, a standard practical guideline of indentifying IVs is worthy of further exploration. For example, Martens et al. have argued that IV analysis may be practically valid if little and moderate confounding exists on the correlation between the IV and the exposure [5]. However, IV assumptions can be easily violated when strong confounding present. Therefore, the standard guideline needs to consider quantifying the level of potential confounding with candidate IVs. Furthermore, a systematic presentation of IV analysis is also critical because it can strengthen argument of a valid IV. For instance, Brookhart et al. have suggested a framework for properly reporting results of IV analysis in comparative safety and effectiveness research [45].

6. Conclusion

In conclusion, use of IV methods is gradually increasing in prescription drug research. We did not find evidence of a dominant IV. Future research should develop standards for identifying candidate IVs and reporting the performance on key IV assumptions.

Key findings

1. Instrumental variable (IV) analysis is gaining popularity in prescription drug research using observational data.
2. Five major types of IVs have been applied to prescription drug research: (i) regional variation, (ii) facility prescribing patterns, (iii) physician preference, (iv) patient history/financial status, (v) calendar time and others.
3. No dominant IV emerged and evidence supporting the validity of IV was often lacking.

What this adds to what was known?

To our knowledge:

1. This is the first systematic review summarizing the use of IVs in prescription drug research.
2. This is the first review to assess the validity of IVs against key assumptions.

What is the implication, what should change now?

1. The five types of IVs summarized in this paper may be helpful for researchers to develop a valid IV using observational data.
2. We recommended that future research develop standards for identifying appropriate IVs and reporting performance on key assumptions.

Acknowledgments

Financial Disclosure: Dr. Briesacher is supported by a Research Scientist Development Award from the National Institute on Aging (K01AG031836) and Dr. Chen is supported by an administrative supplement to the same grant.

References

1. Institute of Medicine. Learning What Works Best: The Natioan's Need for Evidence on Comparativ Effectiveness in Healthcare. 2007 [Janurary 2010]. Available from: http://www.iom.edu/~/media/Files/Activity%20Files/Quality/VSRT/ComparativeEffectivenessWhitePaperESF.pdf
2. Gurwitz JH, Col NF, Avorn J. The exclusion of the elderly and women from clinical trials in acute myocardial infarction. JAMA. 1992 Sep 16; 268(11):1417–22. [PubMed: 1512909]
3. Sorensen HT, Lash TL, Rothman KJ. Beyond randomized controlled trials: a critical comparison of trials with nonrandomized studies. Hepatology. 2006 Nov; 44(5):1075–82. [PubMed: 17058242]
4. Newhouse JP, McClellan M. Econometrics in outcomes research: the use of instrumental variables. Annu Rev Public Health. 1998; 19:17–34. [PubMed: 9611610]
5. Martens EP, Pestman WR, de Boer A, Belitser SV, Klungel OH. Instrumental variables: application and limitations. Epidemiology. 2006 May; 17(3):260–7. [PubMed: 16617274]
6. 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 Dec; 38(6 Pt 1):1423–40. [PubMed: 14727781]
7. Brooks JM, Chrischilles EA, Scott SD, Chen-Hardee SS. Was breast conserving surgery underutilized for early stage breast cancer? Instrumental variables evidence for stage II patients from Iowa. Health Serv Res. 2003 Dec; 38(6 Pt 1):1385–402. [PubMed: 14727779]
8. Earle CC, Tsai JS, Gelber RD, Weinstein MC, Neumann PJ, Weeks JC. Effectiveness of chemotherapy for advanced lung cancer in the elderly: instrumental variable and propensity analysis. J Clin Oncol. 2001 Feb 15; 19(4):1064–70. [PubMed: 11181670]
9. Hadley J, Polsky D, Mandelblatt JS, Mitchell JM, Weeks JC, Wang Q, et al. An exploratory instrumental variable analysis of the outcomes of localized breast cancer treatments in a medicare population. Health Econ. 2003 Mar; 12(3):171–86. [PubMed: 12605463]
10. Leigh JP, Schembri M. Instrumental variables technique: cigarette price provided better estimate of effects of smoking on SF-12. J Clin Epidemiol. 2004 Mar; 57(3):284–93. [PubMed: 15066689]
11. McClellan M, McNeil BJ, Newhouse JP. Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality? Analysis using instrumental variables. JAMA. 1994 Sep 21; 272(11):859–66. [PubMed: 8078163]
12. Permutt T, Hebel JR. Simultaneous-equation estimation in a clinical trial of the effect of smoking on birth weight. Biometrics. 1989 Jun; 45(2):619–22. [PubMed: 2669989]
13. Brookhart MA, Rassen JA, Wang PS, Dormuth C, Mogun H, Schneeweiss S. Evaluating the validity of an instrumental variable study of neuroleptics: can between-physician differences in prescribing patterns be used to estimate treatment effects? Med Care. 2007 Oct; 45(10 Supl 2):S116–22. [PubMed: 17909369]
14. Greene, W. Econometrics analysis. New York: Macmillan; 1990.
15. Brookhart MA, Wang PS, Solomon DH, Schneeweiss S. Evaluating short-term drug effects using a physician-specific prescribing preference as an instrumental variable. Epidemiology. 2006 May; 17(3):268–75. [PubMed: 16617275]
16. J LFB, Silliman RA, Thwin SS, Geiger AM, Buist DS, Prout MN, et al. A most stubborn bias: no adjustment method fully resolves confounding by indication in observational studies. J Clin Epidemiol. 2009 May 19.
17. Dudl RJ, Wang MC, Wong M, Bellows J. Preventing myocardial infarction and stroke with a simplified bundle of cardioprotective medications. Am J Manag Care. 2009 Oct; 15(10):e88–94. [PubMed: 19817511]
18. Groenwold RH, Hak E, Klungel OH, Hoes AW. Instrumental Variables in Influenza Vaccination Studies: Mission Impossible? Value Health. 2009 Aug 20.
19. Lu-Yao GL, Albertsen PC, Moore DF, Shih W, Lin Y, DiPaola RS, et al. Survival following primary androgen deprivation therapy among men with localized prostate cancer. JAMA. 2008 Jul 9; 300(2):173–81. [PubMed: 18612114]

20. Ramirez SP, Albert JM, Blayney MJ, Tentori F, Goodkin DA, Wolfe RA, et al. Rosiglitazone is associated with mortality in chronic hemodialysis patients. J Am Soc Nephrol. 2009 May; 20(5): 1094–101. [PubMed: 19357257]
21. Schneeweiss S, Setoguchi S, Brookhart A, Dormuth C, Wang PS. Risk of death associated with the use of conventional versus atypical antipsychotic drugs among elderly patients. CMAJ. 2007 Feb 27; 176(5):627–32. [PubMed: 17325327]
22. Setoguchi S, Wang PS, Alan Brookhart M, Canning CF, Kaci L, Schneeweiss S. Potential causes of higher mortality in elderly users of conventional and atypical antipsychotic medications. J Am Geriatr Soc. 2008 Sep; 56(9):1644–50. [PubMed: 18691283]
23. Schneeweiss S, Solomon DH, Wang PS, Rassen J, Brookhart MA. Simultaneous assessment of short-term gastrointestinal benefits and cardiovascular risks of selective cyclooxygenase 2 inhibitors and nonselective nonsteroidal antiinflammatory drugs: an instrumental variable analysis. Arthritis Rheum. 2006 Nov; 54(11):3390–8. [PubMed: 17075817]
24. Shetty KD, Vogt WB, Bhattacharya J. Hormone replacement therapy and cardiovascular health in the United States. Med Care. 2009 May; 47(5):600–6. [PubMed: 19319001]
25. Stuart BC, Doshi JA, Terza JV. Assessing the impact of drug use on hospital costs. Health Serv Res. 2009 Feb; 44(1):128–44. [PubMed: 18783453]
26. Wang PS, Schneeweiss S, Avorn J, Fischer MA, Mogun H, Solomon DH, et al. Risk of death in elderly users of conventional vs. atypical antipsychotic medications. N Engl J Med. 2005 Dec 1; 353(22):2335–41. [PubMed: 16319382]
27. Yoo BK, Frick KD. The instrumental variable method to study self-selection mechanism: a case of influenza vaccination. Value Health. 2006 Mar-Apr; 9(2):114–22. [PubMed: 16626415]
28. Zeliadt SB, Potosky AL, Penson DF, Etzioni R. Survival benefit associated with adjuvant androgen deprivation therapy combined with radiotherapy for high- and low-risk patients with nonmetastatic prostate cancer. Int J Radiat Oncol Biol Phys. 2006 Oct 1; 66(2):395–402. [PubMed: 16904843]
29. Costanzo MR, Johannes RS, Pine M, Gupta V, Saltzberg M, Hay J, et al. The safety of intravenous diuretics alone versus diuretics plus parenteral vasoactive therapies in hospitalized patients with acutely decompensated heart failure: a propensity score and instrumental variable analysis using the Acutely Decompensated Heart Failure National Registry (ADHERE) database. Am Heart J. 2007 Aug; 154(2):267–77. [PubMed: 17643575]
30. Tentori F, Albert JM, Young EW, Blayney MJ, Robinson BM, Pisoni RL, et al. The survival advantage for haemodialysis patients taking vitamin D is questioned: findings from the Dialysis Outcomes and Practice Patterns Study. Nephrol Dial Transplant. 2009 Mar; 24(3):963–72. [PubMed: 19028748]
31. Park TR, Brooks JM, Chrischilles EA, Bergus G. Estimating the effect of treatment rate changes when treatment benefits are heterogeneous: antibiotics and otitis media. Value Health. 2008 Mar-Apr; 11(2):304–14. [PubMed: 18380643]
32. Goldman DP, Bao Y. Effective HIV treatment and the employment of HIV(+) adults. Health Serv Res. 2004 Dec; 39(6 Pt 1):1691–712. [PubMed: 15533182]
33. Ikeda N, Gakidou E, Hasegawa T, Murray JG. Understanding the decline of mean systolic blood pressure in Japan: an analysis of pooled data from the National Nutrition Survey, 1986-2002. Bulletin of the World Health Organization. 2008; 86:978–88. [PubMed: 19142299]
34. Salkever D, Slade E, Karakus M. Differential effects of atypical versus typical antipsychotic medication on earnings of schizophrenia patients : estimates from a prospective naturalistic study. Pharmacoeconomics. 2006; 24(2):123–39. [PubMed: 16460134]
35. Salkever DS, Slade EP, Karakus M, Palmer L, Russo PA. Estimation of antipsychotic effects on hospitalization risk in a naturalistic study with selection on unobservables. J Nerv Ment Dis. 2004 Feb; 192(2):119–28. [PubMed: 14770056]
36. Schneeweiss S, Seeger JD, Landon J, Walker AM. Aprotinin during coronary-artery bypass grafting and risk of death. N Engl J Med. 2008 Feb 21; 358(8):771–83. [PubMed: 18287600]
37. Rascati KL, Johnsrud MT, Crismon ML, Lage MJ, Barber BL. Olanzapine versus risperidone in the treatment of schizophrenia : a comparison of costs among Texas Medicaid recipients. Pharmacoeconomics. 2003; 21(10):683–97. [PubMed: 12828491]

38. Zhang Y. Cost-saving effects of olanzapine as long-term treatment for bipolar disorder. J Ment Health Policy Econ. 2008 Sep; 11(3):135–46. [PubMed: 18806303]
39. Cain LE, Cole SR, Greenland S, Brown TT, Chmiel JS, Kingsley L, et al. Effect of highly active antiretroviral therapy on incident AIDS using calendar period as an instrumental variable. Am J Epidemiol. 2009 May 1; 169(9):1124–32. [PubMed: 19318615]
40. Lieberman JA, Stroup TS, McEvoy JP, Swartz MS, Rosenheck RA, Perkins DO, et al. Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. N Engl J Med. 2005 Sep 22; 353(12):1209–23. [PubMed: 16172203]
41. Schneider LS, Tariot PN, Dagerman KS, Davis SM, Hsiao JK, Ismail MS, et al. Effectiveness of atypical antipsychotic drugs in patients with Alzheimer's disease. N Engl J Med. 2006 Oct 12; 355(15):1525–38. [PubMed: 17035647]
42. Hernan MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? Epidemiology. 2006 Jul; 17(4):360–72. [PubMed: 16755261]
43. Grootendorst P. A review of instrumental variables estimation of treatment effects in the applied health sciences. Health Services and Outcomes Research Methodology. 2007; 7(3-4):159–79.
44. Bound J, Jaeger AD, Baker MR. Problems with instrumental variables estimation when the correlation between the instruments and the endougenous exploratory variable is weak. JAMA. 1995; 90:443–50.
45. Brookhart MA, Rassen JA, Schneeweiss S. Instrumental variable methods in comparative safety and effectiveness research. Pharmacoepidemiol Drug Saf. 2010 Mar 30.

Abbreviations

AIDS	Acquired immune deficiency syndrome
ADHF	Acute decompensated heart failure
ADHERE	Acutely Decompensated Heart Failure National Registry
ADT	Adjuvant androgen deprivation
ADAP	AIDS Drug Assistant Program
BVP	Bivariate probit
BOW	Breast Cancer Treatment Effectiveness in Older Women
CATIE	Clinical Antipsychotic Trials of Intervention Effectiveness
COX-2	inhibitor Cyclooxygenase 2 inhibitor
DOPPS	Dialysis Outcomes and Practice Patterns Study
FDA	Food and Drug Administration
GI complications	Gastrointestinal Complications
HCSA	Health care service area
HAART	Highly active antiretroviral therapy
HCSUS	HIV Cost and Services Utilization Study
HIV+	Human immunodeficiency virus positive
IOM	Institute of Medicine
IV	Instrumental variable
MCBS	Medicare Current Beneficiary Survey
NNRTIs	Surveillance, Epidemiology and End Results
NSAID	Non-steroidal anti-inflammatory drug

PACE	Pharmaceutical Assistance Contract for the Elderly
PADT	Primary androgen deprivation therapy
RCT	Randomized controlled trial
SCAP	Schizophrenia Care and Assessment Program
SEER	Surveillance, Epidemiology and End Results
3SLS	Three stage least square
2SLS	Two stage least square
2SRI	Two stage residual inclusion
VD	Vitamin D

Figure 1. Instrumental variable analysis

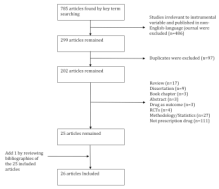

Figure 2. Literature searching strategy

J Clin Epidemiol. Author manuscript; available in PMC 2012 June 1.

Table 1

Description of studies arranged by publication year

Year	Author	Data source	Sample size	Age	Sex
2001	Earle et al. [8]	Surveillance, Epidemiology and End Results (SEER) Tumor Registry	6,232	65+	Both
2003	Rascati et al. [37]	Texas Medicaid population	2,885	21-66	Both
2004	Goldman et al. [32]	HIV Cost and Services Utilization Study (HCSUS)	2,013	24-49	Both
2004	Salkever et al. [35]	Schizophrenia Care and Assessment Program (SCAP)	937	18+	Both
2005	Wang et al. [26]	PACE	22,890	65+	Both
2006	Salkever et al. [34]	SCAP	336	18+	Both
2006	Brookhart et al. [15]	Pharmaceutical Assistance Contract for the Elderly (PACE)	49,919	65+	Both
2006	Schneeweiss et al. [23]	PACE	37,650	65+	Both
2006	Yoo et al. [27]	Medicare Current Beneficiary Survey (MCBS)	4,338	65+	Both
2006	Zeliadt et al. [28]	SEER Tumor Registry	31,643	65-85	Male
2007	Costanzo et al. [29]	Acutely Decompensated Heart Failure National Registry (ADHERE) data base	99,963	Average 72.6	Both
2007	Schneeweiss et al. [21]	Population from British Columbia	37,241	65+	Both
2008	Ikeda et al. [33]	Nationally representative of Japan	192,457	20+	Both
2008	Lu-Yao et al. [19]	SEER Tumor Registry	19,271	74-79	Male
2008	Park et al. [31]	IOWA Medicaid population	36,585	0-12	Both
2008	Schneeweiss et al. [36]	Premier Perspective Comparative Database	25,784	18+	Both
2008	Setoguchi et al. [22]	Population from British Columbia	37,241	65+	Both
2008	Zhang et al. [38]	Marketscan Data (private pharmaceutical claims data)	933	17-64	Both
2009	Bosco et al. [16]	Breast Cancer Treatment Effectiveness in Older Women (BOW) cohort	539	65+	Female
2009	Cain et al. [39]	Multicenter AIDS cohort	614	Average 35	Male
2009	Dudl et al. [17]	Kaiser Permanente	170,024	55+	Both
2009	Groenwold et al. [18]	Computerized Utrecht General Practitioner Database	60,000	65+	Both
2009	Ramirez et al. [20]	Dialysis Outcomes and Practice Patterns Study (DOPPS)	2,393	62+	Both
2009	Shetty et al. [24]	Health care utilization project's nationwide inpatient sample	Nationally representative sample (No exact sample size)	65+	Both
2009	Stuart et al. [25]	Medicare Current Beneficiary Survey (MCBS)-Medicare population	3,101	65+	Both
2009	Tentori et al. [30]	DOPPS	38,006	Average 60	Both

Table 2

Table 2a: IV, exposure and outcome (arrange studies by IV type)					
IV Type	**Author**	**IV**	**Exposure**	**Outcome**	**Outcome Type**
Regional variation	Earle et al. [8]	Prevalence of chemotherapy in HCSA. Grouped HCSAs into quintiles based on the prevalence.	Chemotherapy for lung cancer	Mortality	Effectiveness
	Goldman et al. [32]	3 prescriptions/month by Medicaid in a state in 1997	HAART	Returning to work	Effectiveness
		No coverage of Non-Nucleoside Analogue Reverse Transcriptase Inhibitors (NNRTIs) by the state AIDS Drug Assistant Program (ADAP) in 1997		Remaining employed	
				Hours of work	
	Ikeda et al. [33]	Proportion of people with hypertension who were on treatment in prefecture of residence	Antihypertensive drugs	Systolic blood pressure (SBP)	Effectiveness
	Lu-Yao et al. [19]	Proportion of patients who received PADT in each HSA	PADT	Mortality	Effectiveness
	Park et al. [31]	Physician per capita in the 10-mile radius around patient zip code	Antibiotics	Cure of Otitis media (OM)	Effectiveness
	Salkever et al. [34]	Percentage of Medicaid antipsychotic treatment for atypicals in the state in baseline quarter	Atypical antipsychotics	Earnings from work	Effectiveness
	Salkever et al. [35]	Percentage of Medicaid antipsychotic treatment for atypicals in the state in baseline quarter	Atypical antipsychotics	Hospitalizations	Effectiveness
	Zeliadt et al. [28]	Proportion of people treated with Adjuvant Androgen Deprivation Therapy (ADT) in HCSA	ADT	Mortality	Effectiveness
Facility prescribing patterns	Dudl et al. [17]	Facility proportion of patients on bundle of cardioprotective drugs	Bundle of cardioprotective drugs	Hospitalization of MI or stroke	Effectiveness
	Ramirez et al. [20]	Facility proportion of patients on Rosiglitazone	Rosiglitazone	Cardiovascular hospitalization and all cause mortality	Adverse event
	Salkever et al. [34]	Provider was a Veteran Administration (VA) medical center	Atypical antipsychotics	Earnings from work	Effectiveness
		Provider was a community mental health center with no teaching affiliation			
		Provider reported date when atypicals were added to formulary			
	Salkever et al. [35]	Provider was a VA medical center	Atypical antipsychotics	Hospitalizations	Effectiveness
		Provider was a community mental health center with no teaching affiliation			
		Provider reported date when atypicals were added to formulary			

Table 2a: IV, exposure and outcome (arrange studies by IV type)					
IV Type	**Author**	**IV**	**Exposure**	**Outcome**	**Outcome Type**
	Tentori et al. [30]	Percentage of patients at a facility receiving vitamin D (VD)	VD	Mortality	Adverse event
		Adjusted percentage of patients in a facility receiving VD			

Table 2b: IV, exposure and outcome (arrange studies by IV type)					
IV Type	**Author**	**IV**	**Exposure**	**Outcome**	**Outcome Type**
Physician preference	Bosco et al. [16]	Patient's surgeon's chronologically preceding patient's receipt of adjuvant chemotherapy	Adjuvant chemotherapy	Breast cancer recurrence	Effectiveness
	Brookhart et al. [15]	Physician's last prescription drug was a COX-2 or a non-selective NSAID	COX-2 or Non-selective NSAID	GI complications	Adverse event
	Groenwold et al. [18]	General practitioner group's specific vaccination rate	Influenza vaccine	Mortality	Effectiveness
	Schneeweiss et al. [36]	Proportion of patients who were administered Aprotinin by a surgeon	Aprotinin	Mortality	Adverse event
	Schneeweiss et al. [21]	Physician's most recent antipsychotic prescription	Atypical antipsychotics	Mortality	Adverse event
	Schneeweiss et al. [23]	Physician's most recent NSAID prescription	Selective or non-selective NSAIDs	GI Complications	Adverse event
	Setoguchi et al. [22]	Physician's most recent antipsychotic prescription	Atypical antipsychotics	Mortality	Adverse event
	Wang et al. [26]	Physician's most recent antipsychotic prescription	Atypical antipsychotics	Mortality	Adverse event
Patient history/financial status	Groenwold et al. [18]	History of gout	Influenza vaccine	Mortality	Effectiveness
		History of orthopedic morbidity			
		History of antacid medication			
	Stuart et al. [25]	Drug coverage	Prescription drug counts	Hospitalizations; Total hospital expenditures	Effectiveness
	Yoo et al. [27]	History of gout or arthritis	Influenza vaccine	Influenza related hospitalization or death	Effectiveness
Calendar time	Cain et al. [39]	Calendar period (before and after 1996 when HAART was available)	HAART	Clinical AIDS	Effectiveness
	Rascati et al. [37]	Time in months between initiation of treatment with an antipsychotic and Oct 1996 when atypical agent was first available for purchasing.	Atypical antipsychotics	Total hospital cost and Schizophrenia-related cost	Effectiveness

Table 2b: IV, exposure and outcome (arrange studies by IV type)					
IV Type	**Author**	**IV**	**Exposure**	**Outcome**	**Outcome Type**
	Shetty et al. [24]	Release of Women's Health Initiative (WHI) data in July of 2002	Hormone replacement therapy	Cardiovascular outcomes	Adverse event
	Zhang et al. [38]	Time of the Food and Drug Administration's (FDA) approval of olanzapine	Olanzapine	Non-drug spending	Effectiveness
				Total services spending	
				Mental health services spending	
Other(Statistical)	Costanzo et al. [29]	Patient's propensity to receive treatment (Propensity score)	Intravenous therapies for acute decompensated heart failure (ADHF)	Mortality	Effectiveness

Table 3

Table 3a: Analytical approach and validation of two assumptions for IV (studies are arranged by quality score)					
Author	**Analytical approach**	**Assumption 1: IV is associated with treatment assignment**	**Assumption 2: IV is not associated with measured and unmeasured confounders and is only associated with the outcome through treatment assignment**	**Quality of IV**	
				Score	**Reason for the score**
Bosco et al. [16]	2 stage model: **1** First step: using logistic regression to estimate the probability of receiving treatment. **2** Second step: using survival analysis to estimated the risk of breast cancer recurrence by including the probability of receiving treatment from the first step	They used linear regression to test the strength of association between IV and treatment assignment (coefficient=23.7%) and reported that the strength of IV was comparable to strength from another study using similar IV (coefficient = 23%).	IV reduced imbalance for some patient characteristics and increased the imbalance for others. This implied that IV was still associated with measured and unmeasured confounders.	1	Did not meet the second assumption
Brookhart et al. [15]	2 stage least square (2SLS)	The authors reported difference in treatment assignment between the two levels of the IV. For physician whose last prescription was COX-2, the probability that the physician's next prescription would be COX-2 was 77%. For physician whose last prescription was non-selective NSAID, the probability was 55%.	IV was weakly associated with observed characteristics implying that it might be associated with unobservables. The authors mentioned that the IV might be associated with the outcome through other paths besides treatment and acknowledged this limitation.	1	Did not meet the second assumption
Cain et al. [39]	**1** Point estimates for both probability of treatment assignment and outcome were computed by pooled logistic regression. **2** Bootstrapping sample 500 times was used to estimate confidence interval.	Compare difference in treatment rate before and after the calendar period. They also stated that this assumption was well documented.	They used inverse probability of calendar period weighs to relax the third assumption.	1	Did not assess the second assumption
Costanzo et al. [29]	2 stage IV analysis, bootstrapping for confidence interval	Propensity score was highly associated with treatment assignment.	Measured confounders should be balanced among patients with similar propensity score.	1	Propensity score only balanced measured confounders

Table 3b: Analytical approach and validation of two assumptions for IV (studies are arranged by quality score)					
Dudl et al. [17]	Not abstracted due to unclear description of method	Facility use rates from 32% in the lowest-using quintile of facilities to 49.1% in the highest using quintile	IV reduced imbalance of patient characteristics. However, unmeasured facility-level confounder might still bias the results.	1	Did not meet the second assumption

Groenwold et al. [18]	**1** Linear regression to estimate the difference in outcome and difference in vaccination rate between the two IV categories. **2** Bootstrapping for confidence interval	History of gout was associated with vaccination (OR=1.56; 95%CI 1.23-1.97) History of orthopedic morbidity was associated with vaccination (OR=1.16; 95CI 1.10-1.22) History of antacid medication was associated with vaccination (RD=18.1%; 95CI 6.1%-10.1%) Vaccination rates among GP group practices ranged from 68.1-77.9%	All 4 IVs were significantly associated with patient characteristics indicating weak IVs.	1	Did not meet the second assumption
Rascati et al. [37]	2 stage model: **1** First stage: using probit regression to estimate the probability of receiving treatment **2** Second stage: using OLS regression to estimate the hospital cost or total schizo-related cost	In the first stage the IV was significantly associated with treatment assignment (Marginal effect = -0.0074; P<0.0001)	NA	1	No discussion of the second assumption
Setoguchi et al. [22]	2SLS	Instrument was associated with treatment assignment (OR=6.1 95% CI 5.8-6.4)	NA	1	No discussion of the second assumption
Yoo et al. [27]	BVP/2SLS reported BVP results	They stated that they examined F-statistic using pseudo-R square in BVP model. But no detailed data were reported.	NA	1	No discussion of the second assumption

Table 3c: Analytical approach and validation of two assumptions for IV (studies arranged by quality score)

Earle et al. [8]	**1** Difference between survival in the highest and the lowest quintiles of HCSAs/difference in the probability of receiving treatment between the two **2** Bootstrapping for confidence interval	HCSA was highly correlated with the likelihood of receiving chemotherapy.(F statistic= 7792, R square = 0.71)	They compared the observable characteristics between HCSAs with the highest and lowest prevalence of chemotherapy and found the IV reduced imbalance of observed characteristics.	2	Assessed both assumptions
Goldman et al. [32]	Bivariate probit (BVP) 2SLS	The authors showed that the IV predicted the use of HAART. Average predicted probability of HAART receipt in return to work sample ranged from 31% in non-coverage group to 38% in coverage group. It ranged from 33% to 46% in the remaining employed sample and ranged from 34% to 44% in the hours of work sample.	They argued that patients have little influence on state policy, so state policy instrument is exogenous.	2	Assessed the first assumption and discussed the rationale for the second assumption
Ikeda et al. [33]	2SLS	The authors reported there was an association between the IV and treatment assignment (OR for IV = 47.7)	The authors reported that there was no significant difference in SBP among regions with different treatment rates.	2	Assessed both assumptions

Lu-Yao et al. [19]	**1** Difference between survival in highest and the lowest quintiles of HSAs/ difference in the probability of receiving treatment between the two **2** Bootstrapping for confidence interval	The authors reported that PADT use varied by HSAs. (31%-53%)	They presented data showing that measured confounders were balanced by IV.	2	Assessed both assumptions
Park et al. [31]	2SLS	They provided Chow F-statistics in the first stage model and reported that IV was significantly associated with treatment assignment. (F=11.98, P<0.001)	They reported that imbalance was reduced by IV in some important confounders such as severity and comorbidity. But it was increased in other variables such as race. They used Hausman test statistic to examine that IV had no direct and indirect effect on the outcome through an unmeasured confounder.	2	Assessed both assumptions

Table 3d: Analytical approach and validation of two assumptions for IV (studies arranged by quality score)

Ramirez et al. [20]	2 stage model: **1** First stage: using linear mixed model to estimate the probability of receiving rosiglitazone **2** Second stage: survival analysis to estimate the risk of cardiovascular hospitalization by including probability of treatment from the first stage	They reported that IV explained the variation of treatment assignment. (R square = 0.166 for facility and case-mix variable; R square = 0.031 for case-mix variable alone.	Patient characteristics were balanced by the IV. (Table 3)	2	Assessed both assumptions
Salkever et al. [34]	2 stage generalized least square	Four IVs were significantly associated with treatment assignment in the first stage. Total R square was 0.29.	Sargan test for over-identification.	2	Assessed both assumptions
Salkever et al. [35]	BVP	IVs were significantly associated with treatment assignment for patient younger than 45yrs (P<0.05 for 3 out of 4 IVs) but not for patients older than 44yrs.	Including the instruments in the hospitalization regression and compute relevant test statistics based on likelihood ratios	2	Assessed both assumptions
Schneeweiss et al. [36]	2SLS	They reported that aprotinin utilization varied by surgeons	They argued that patients were not likely to choose their surgeon on the basis of the surgeon's preference for a specific antifibrinolytic agent implying patient characteristics would be independent of surgeon preference.	2	Assessed both assumptions
Schneeweiss et al. [21]	2SLS	Instrument was associated with treatment assignment (OR=6.1 95% CI 5.8-6.4)	Imbalance of study variables was reduced by IV.	2	Assessed both assumptions
Schneeweiss et al. [23]	2SLS	Physicians with different preference of NSAID varied by their COX-2 inhibitor utilizations.	Imbalance of study variables was reduced by IV.	2	Assessed both assumptions

Shetty et al. [24]	3 stage least square (3SLS)	They provided argument that release of the WHI data affected the use of HRT.	In the limitation section, they discussed that unmeasured confounders were not likely to affect the results.	2	Assessed both assumptions
Stuart et al. [25]	2 stage residual inclusion (2SRI)	They reported the drug coverage was significantly associated with treatment assignment. (t=6.993, p <0.001)	Residual derived from first stage was significant in the second stage indicating regression model biased and 2SRI was unbiased by unmeasured confounders.	2	Assessed both assumptions

Table 3e: Analytical approach and validation of two assumptions for IV (studies arranged by quality score)

Tentori et al. [30]	A facilities propensity to prescribe VD was modeled as the adjusted proportion of patients on VD at a facility, estimated by linear mixed regression. And 2 IVs were further adjusted in Cox model to predict patient's risk of mortality.	Propensity to prescribe was associated with VD.	They reported that imbalance of patient characteristics was reduced by IV.	2	Assessed both assumptions
Wang et al. [26]	2sls	They argued that antipsychotic prescriptions varied by physicians. But no data were reported.	**1** For confounder between the treatment and the outcome: They reported that imbalanced of measured confounders was reduced by IV. This was validated in another study (ref 4). **2** For confounder between the IV and the outcome: Exclusion criteria were tested by examining the association between physician preference of antipsychotics and hazard concomitant prescription of Tricycle antidepressant (TCA) and long-acting benzodiazepine. This was validated in another study (ref 4) There was no association when they restricted ample to primary physicians.	2	Assessed both assumptions
Zeliadt et al. [28]	2SRI	Reported change of the use of olanzapine before and after March 2000 and reported the first stage F-statistic = 27.63 (P<0.05)	Discussed that timing of approval is not associated with physician behavior patterns or other unmeasured variables.	2	Assessed both assumptions
Zhang et al. [38]	Difference between survival in highest and the lowest HCSAs/difference in the proportion of receiving treatment between the two was used as iV estimate. And they used bootstrapping 1000 samples to estimate confidence interval.	They used linear model to estimate the association between IV and treatment assignment and reported F-statistic was greater than 10.	Imbalance of patient characteristics was reduced by IV.	2	Assessed both assumptions