Unmeasured confounding IV PERR SCCS Observational studies
PROTECTCausal Inference Statistical methods পরিসংখ্যানগত পদ্ধতি
Statistische methoden IV PERR SCCS Pseudo Randomization
Pharmacoepidemiology IV Analyses PERR SCCS
General Practice Databases Residual Confounding
Electronic Healthcare Databases Epidemiology
Time-fixed Confounding Time-fixed Exposure
Time-varying Exposure
Time-varying Confounding
Non-randomized Studies
IV PERR SCCS
PROTECT 2SLS
IV IV

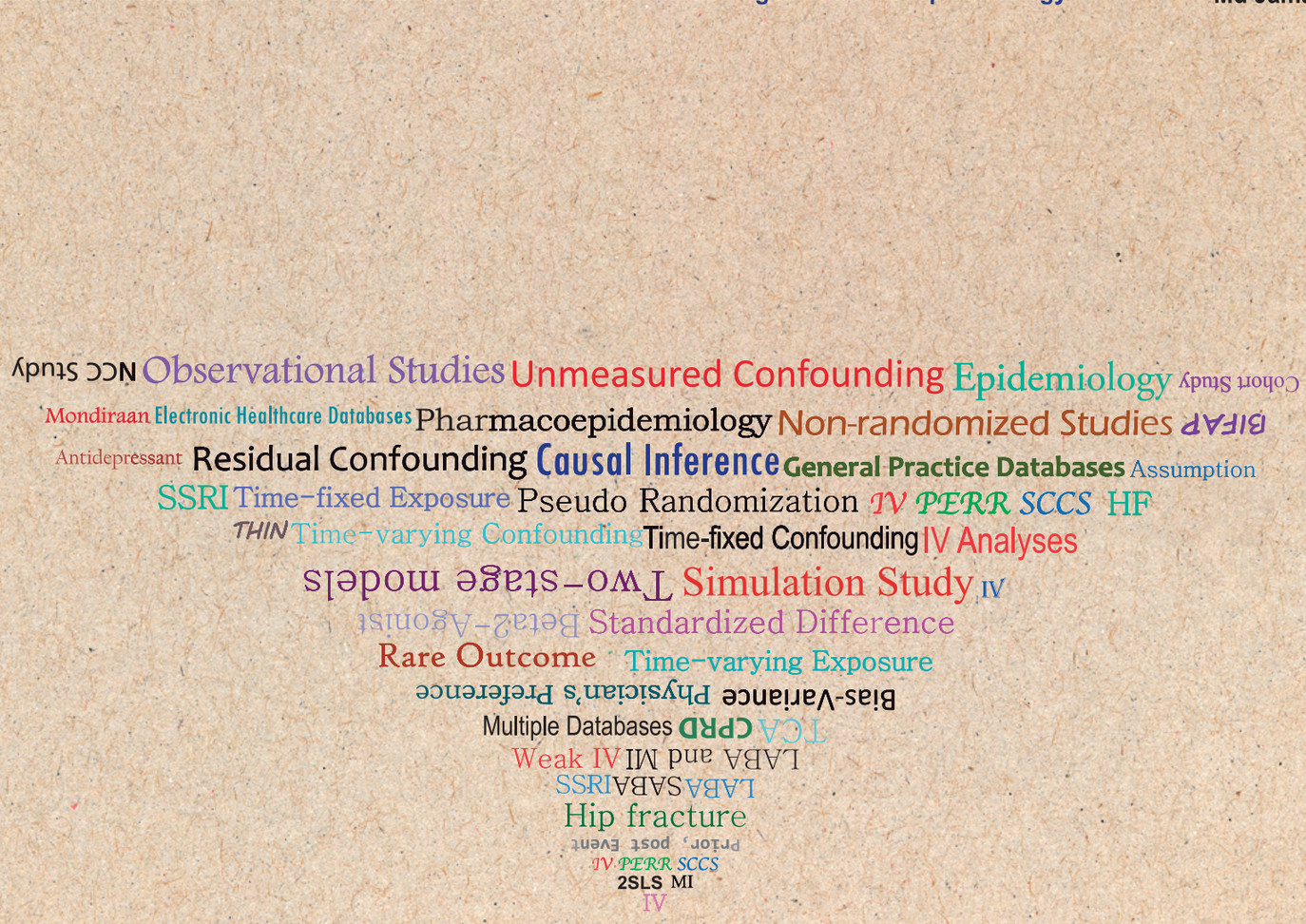
PERFORMANCE OF STATISTICAL METHODS TO CONTROL FOR UNMEASURED CONFOUNDING IN PHARMACOEPIDEMIOLOGY

Focus on Instrumental Variable Analysis

Md Jamal Uddin

Performance of Statistical Methods to Control for Unmeasured Confounding in Pharmacoepidemiology

Md Jamal Uddin



PERFORMANCE OF STATISTICAL METHODS TO CONTROL FOR UNMEASURED CONFOUNDING IN PHARMACOEPIDEMIOLOGY

FOCUS ON INSTRUMENTAL VARIABLE ANALYSIS

Md Jamal Uddin

Division of Pharmacoepidemiology and Clinical Pharmacology
Utrecht Institute for Pharmaceutical Sciences
University of Utrecht, Utrecht, the Netherlands

The work presented in this thesis was performed at the Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht, the Netherlands.

The research presented in this thesis was conducted as part of the PROTECT consortium (Pharmacoepidemiological Research on Outcomes of Therapeutics by a European ConsorTium) which is a public-private partnership coordinated by the European Medicines Agency. The PROTECT project is supported by the Innovative Medicine Initiative Joint Undertaking (www.imi.europa.eu) under Grant Agreement no 115004, resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and EFPIA companies' in kind contribution, the division of Pharmacoepidemiology & Clinical Pharmacology, Utrecht University, also received a direct financial contribution from Pfizer.

Financial support for printing of this thesis was kindly supported by Utrecht Institute for Pharmaceutical Sciences (UIPS) and Het Nederlands Bijwerkingen Fonds (NBF). Moreover, financial support by the Dutch Heart Foundation for the publication of this thesis is gratefully acknowledged. In addition, I am grateful for the financial support by ChipSoft.

Layout: Off Page

Cover Design: Md Jamal Uddin

Printed by: Off Page

Uddin, Md Jamal

Performance of statistical methods to control for unmeasured confounding in pharmacoepidemiology. Focus on instrumental variable analysis

ISBN/EAN: 978-94-6182-515-5

PERFORMANCE OF STATISTICAL METHODS TO CONTROL FOR UNMEASURED CONFOUNDING IN PHARMACOEPIDEMIOLOGY

FOCUS ON INSTRUMENTAL VARIABLE ANALYSIS

Validiteit en toepasbaarheid van statistische methoden ter correctie van ongemeten confounding in farmaco-epidemiologisch onderzoek

Focus op instrumentele variabele analyse

(met een samenvatting in het Nederlands)

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof. dr. G.J. van der Zwaan, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op maandag 15 december 2014 des ochtends te 10.30 uur

door

Md Jamal Uddin

geboren op 1 augustus 1977 te Comilla, Bangladesh

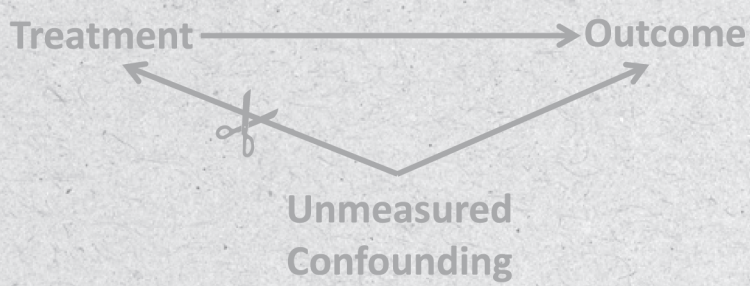
Promotoren:	Prof. dr. A. de Boer Prof. dr. K.C.B. Roes
Copromotoren:	Dr. R.H.H. Groenwold Dr. O.H. Klungel

To my Parents, wife Jannat, and daughter Nudrat

TABLE OF CONTENTS

Unmeasured confounding IV PERR SCCS Observational studies
PROTECT Causal Inference Statistical methods পরিসংখ্যানগত পদ্ধতি
Statistische methoden IV PERR SCCS Pseudo Randomization
Pharmacoepidemiology IV Analyses PERR SCCS
General Practice Databases Residual Confounding
Electronic Healthcare Databases Epidemiology
Time-fixed Confounding Time-fixed Exposure
Time-varying Exposure
Time-varying Confounding
Non-randomized Studies
IV PERR SCCS
PROTECT 2SLS
IV IV

CHAPTER I
GENERAL INTRODUCTION



In medical and health sciences research, randomized controlled trials are considered the gold standard to estimate the causal effect of treatments.[1–3] The major characteristic of the randomized trial is the control, by the investigator, over treatment assignment by means of randomization. However, randomized trials may be of limited use or not feasible in several situations, for instance, in the case of rare adverse drug events, or when the outcomes of interest are far ahead in the future (e.g., long term consequences of statins use and the risk of cancer).[1–5] Also, in some situations, randomized trials are too expensive to conduct or may not be ethical (e.g., to investigate new suspected adverse events of existing treatments, such as selective serotonin reuptake inhibitor use and the risk of hip fracture).[1–5] In such situations, observational (non-randomized) studies may be viable alternatives to provide important evidence on the comparative safety and effectiveness of different treatments.[1,2,5–8]

In observational studies, the assignment of subjects into a treated group versus a control group is outside the control of the investigator as in daily clinical practice treatment assignment is generally based on the physician's perception of the patients risk of a particular outcome.[6,9] Because of lack of randomization, prognostic patient characteristics are typically unevenly distributed among the study groups. Hence, a direct comparison between treatment groups will possibly be biased due to confounding.[6,9–11]

A confounder is a prognostic factor of the outcome which is associated with the exposure and causes a biased estimate of the exposure effect on the outcome.[12,13] For example in Figure 1, the variable pre-treatment blood pressure is associated with the treatment, ACE-inhibitor: angiotensin-converting-enzyme inhibitor, and is an independent risk factor of the

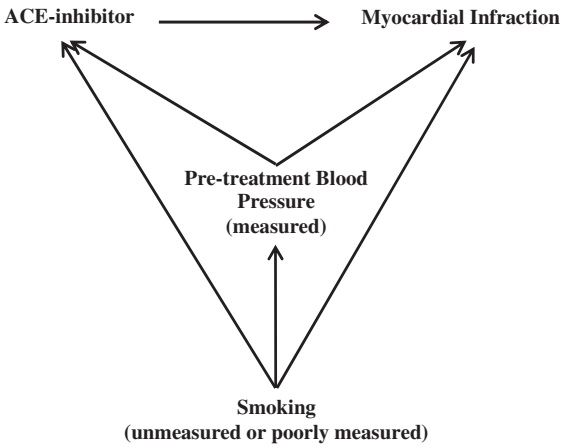


Figure 1. Directed acyclic graph of observational studies to illustrate of the concept of the confounding

Treatment/Exposure: ACE-inhibitor: angiotensin-converting-enzyme inhibitor, Outcome: Myocardial Infraction, Measured confounder: Pre-treatment Blood Pressure, Unmeasured/poorly measured confounder: Smoking.

outcome, myocardial infarction. Several statistical methods such as restriction, matching, stratification, multivariable regression, propensity score, inverse probability weightings, can control for confounding by measured confounders (like pre-treatment blood pressure) in observational studies.[9,11,14–16]

There may also be unmeasured or poorly measured prognostic factors (e.g., smoking) of the outcome that are also associated with the exposure, these are referred to as unmeasured confounders.[12] However, the aforementioned methods can only control for confounding by *measured* confounders. Hence, confounding by unmeasured patient characteristics may impair the validity of the study results.[9,11,14–16] Nevertheless, alternative statistical methods have been proposed to detect or to control for unmeasured confounding either in the design or analysis of an observational study.[9,16–29] Here, we focus on three methods that were recently developed or that have been increasingly used over the last decade to control for unmeasured confounding: instrumental variable (IV) analysis, prior event rate ratio (PERR) adjustment method, and self-controlled case series (SCCS) design.[20,21,24,25,30–35]

INSTRUMENTAL VARIABLE ANALYSIS

IV analysis is one of the popular methods that has primarily been used in econometrics and social sciences, but has appeared in epidemiologic research over the last decade to control for unmeasured confounding.[7,33,34,36] In observational studies, IV analysis tries to mimic a randomized study, in which treatment assignment is related to the actual treatment received and treatment assignment only affects outcome through the received treatment (hence, the term pseudo-randomisation that is sometimes used for IV analysis).[30] As an IV is assumed to affect the outcome only through the treatment/exposure under study, IV analysis potentially controls for both measured and unmeasured confounding because it implies that all measured and unmeasured confounders could be equally distributed among the exposure groups made by the IV (similar to a randomized controlled trial). This means that an IV should satisfy three key assumptions: 1) the IV is associated with the treatment/exposure under study, 2) the IV affects the outcome only through the exposure (exclusion restriction), and 3) the IV is independent of confounders.[30–32]

An example of an IV in an observational study is illustrated in Figure 2.[37] The idea behind this IV, physician prescribing preference, is that physicians differ with respect to their preference for conventional versus atypical antipsychotic medication: to a similar patient one physician will prescribe conventional antipsychotic medication, while another physician will prescribe an atypical antipsychotic medication. Furthermore, it is assumed that the preference is only influencing mortality (the outcome) through conventional or atypical antipsychotic medication exposure and is independent of patient characteristics.[37]

In 2006, Brookhart et al.[20] first applied IV analysis using physician prescribing preference as an IV in a large healthcare database study to assess the adverse effects of drugs and since then IV analysis has increasingly been used in pharmacoepidemiology.[33,34,38] Nevertheless,

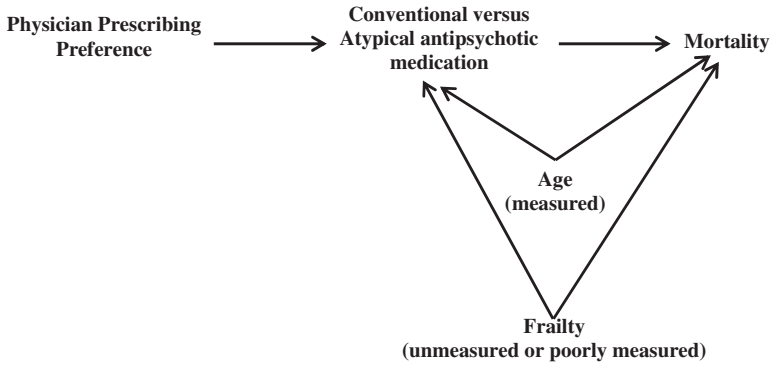


Figure 2. Directed acyclic graph of observational studies to illustrate of the concept of the instrumental variable

Instrumental variable: Physician prescribing preference, Treatment/Exposure: Conventional versus atypical antipsychotic medication, Outcome: Mortality, Measured confounder: Age, Unmeasured/ poorly measured confounder: Frailty.

the assumptions of IV analysis may not always be met and thus the validity of the results of IV analysis are debated.[31,33,38–40]

In recent years, there have been many studies for evaluating the methodological aspects of IV analysis.[31,32,36,41–50] However, there are still many situations where the method has not been studied well or the performance of the method remains unclear. For example, although the application of IV analysis in cohort studies is relatively well understood,[20,39,51–54] the application and performance of IV methodology in the (nested) case-control design is less clear. Another example is the evaluation of the assumptions underlying IV analysis, e.g. methods are available to verify the first IV assumption empirically,[32,55–58] but there are no well-established methods to test the other assumptions. In addition, multiple database studies of drug effects from different countries have widely been conducted,[59–61] yet comparisons across databases are scarce.

PRIOR EVENT RATE RATIO ADJUSTMENT METHOD

The PERR adjustment method was recently developed and validated within electronic health records databases.[21,62] This method theoretically controls for all time-fixed measured and unmeasured confounding.[17] It can be applied in a setting where periods can be identified during which none of the study participants were treated, yet after some moment in time in part of the study population treatment was initiated.[21,62] The exposure effect is then estimated by the ratio of two rate ratios: the rate ratio after initiation of exposure and the rate ratio prior to initiation of exposure.[17,63] It requires assumptions about confounding effects being constant over time, absence of confounder-by-treatment interaction, and non-lethal outcomes.[16,62,63] Only one previous study by Yu et al.[64] evaluated and addressed

some critical methodological issues of this method. Therefore, further methodological investigations are required to fully understand this method and provide recommendations for proper application in pharmacoepidemiology.

SELF-CONTROLLED CASE SERIES DESIGN

The SCCS design was originally developed to evaluate vaccine safety and has been increasingly used in pharmacoepidemiological studies using healthcare databases.[24,35,65] The SCCS uses information from case subjects only and is thus similar to the case-crossover design (subjects act as their own controls). This design has been used to evaluate the association between a transient exposure and an acute event.[25,65] The key assumptions of this design are: 1) events are recurrent and independent or are unique and uncommon over the study periods; 2) the occurrence of an event must not alter the probability of subsequent exposure; and 3) the occurrence of the event should not censor the observation period.[24,25,65] Although several studies[25,65–69] acknowledged that violations of the key SCCS assumptions may bias exposure effects, little is known about the impact of such violations in empirical data.

OBJECTIVES

The overall aim of this thesis is to evaluate the performance of statistical methods to control for unmeasured confounding and to provide recommendations for proper application in pharmacoepidemiology.

OUTLINE OF THE THESIS

Chapter 2 contains two introductory studies of IV analysis. In **chapter 2.1**, we describe the conceptual framework of IV analysis. We also discuss why application of IV analysis in pharmacoepidemiology may be hampered. In **chapter** 2.2, we review and describe IV methods for continuous as well as binary outcomes, exposures, and IVs and highlight the strengths and limitations of the different methods for IV analysis.

Chapter 3 contains two simulation studies of IV analysis. In **chapter 3.1**, extensive simulation studies are described that evaluate IV analysis for different combinations of continuous or binary IV, exposure, and outcome in both the cohort and the nested case-control design. C**hapter 3.2** describes a simulation study to assess the performance of balance measures commonly used in propensity score methods (i.e., standardized difference) to quantitatively falsify the 3rd assumption of IV analysis. Additionally, this measure is applied in an empirical study of the relation between beta2-agonist use and myocardial infarction.

Chapter 4 contains two empirical studies of IV analysis in which exposures of interest are time-fixed as well as time-varying. In **chapter 4.1**, different physician's prescribing preference based IVs are evaluated in two primary care databases (the UK-based Clinical Practice

Research Datalink and the Dutch Mondriaan database) in a study of inhaled long-acting beta2-agonist use and the risk of myocardial infarction. In **chapter 4.2**, the performance of IV analysis is assessed using different IVs in multiple databases in a study of antidepressant use and hip fracture. The databases used for this study are the health improvement network (THIN, UK); the Spanish Base de datos para la Investigación Farmacoepidemiológica en Atención Primaria (BIFAP; Spain); and the Dutch Mondriaan general practice database.

Chapter 5 includes two studies in which PERR adjustment and the SCCS design are evaluated, respectively. In **chapter 5.1,** we describe a simulation study to investigate the PERR adjustment method to clarify constraints and to understand its proper applicability in pharmacoepidemiology. In **chapter 5.2**, the impact of violation of the assumptions underlying the SCCS design is assessed in a study of antidepressants use and risk of hip fracture using data from two general practice databases (Mondriaan, Netherlands and THIN, UK).

Chapter 6 contains a general discussion in which a comparison is made between the different methods and recommendations are provided for choosing between the different methods.

REFERENCES

1. Grobbee DE, Hoes AW. Clinical epidemiology: principles, methods, and applications for clinical research. Jones & Bartlett Learning; 2009.
2. Vandenbroucke JP. When are observational studies as credible as randomised trials? *The Lancet* 2004; 363(9422): 1728-31.
3. Berger ML, Martin BC, Husereau D, et al. A Questionnaire to Assess the Relevance and Credibility of Observational Studies to Inform Health Care Decision Making: An ISPOR-AMCP-NPC Good Practice Task Force Report. *Value in Health* 2014; 17(2): 143-56. doi:http://dx.doi.org/10.1016/j.jval.2013.12.011.
4. Feinstein AR. Current problems and future challenges in randomized clinical trials. *Circulation* 1984; 70(5): 767-74.
5. Black N. Why we need observational studies to evaluate the effectiveness of health care. *Br Med J* 1996; 312(7040): 1215-8.
6. Brookhart MA, Stürmer T, Glynn RJ, Rassen J, Schneeweiss S. Confounding control in healthcare database research: Challenges and potential approaches. *Med Care* 2010; 48(6 SUPPL.): S114-20.
7. Baiocchi M, Cheng J, Small DS. Instrumental variable methods for causal inference. *Stat Med* 2014; 15(33(13)): 2297-340. doi:10.1002/sim.6128.
8. McMahon AD. Approaches to combat with confounding by indication in observational studies of intended drug effects. *Pharmacoepidemiol Drug Saf* 2003; 12(7): 551-8.
9. Klungel OH, Martens EP, Psaty BM, et al. Methods to assess intended effects of drug treatment in observational studies are reviewed. *J Clin Epidemiol* 2004; 57(12): 1223-31.
10. Bosco JLF, Silliman RA, Thwin SS, et al. A most stubborn bias: no adjustment method fully resolves confounding by indication in observational studies. *J Clin Epidemiol* 2010; 63(1): 64-74. doi:http://dx.doi.org/10.1016/j.jclinepi.2009.03.001.
11. Groenwold RHH, Hak E, Hoes AW. Quantitative assessment of unobserved confounding is mandatory in nonrandomized intervention studies. *J Clin Epidemiol* 2009; 62(1): 22-8.

12. Velentgas P, Dreyer NA, Nourjah P, Smith SR, Torchia MM, eds. Developing a Protocol for Observational Comparative Effectiveness Research: A User's Guide. AHRQ Publication No. 12(13)-EHC099. Rockville, MD: Agency for Healthcare Research and Quality; January 2013. www.effectivehealthcare.ahrq.gov/Methods-OCER.cfm.
13. VanderWeele TJ, Shpitser I. On the definition of a confounder. *The Annals of Statistics* 2013; 41(1): 196-220.
14. Robins JM, Hernán MÁ, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology* 2000; 11(5): 550-60.
15. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983; 70(1): 41-55.
16. Alemayehu D, Alvir JM, Jones B, Willke RJ. Statistical issues with the analysis of nonrandomized studies in comparative effectiveness research. *Journal of managed care pharmacy : JMCP* 2011; 17(9 Suppl A): S22-26.
17. Schneeweiss S, Seeger JD, Smith SR. Methods for developing and analyzing clinically rich data for patient-centered outcomes research: An overview. *Pharmacoepidemiol Drug Saf* 2012; 21(SUPPL.2): 1-5.
18. Groenwold RHH, Hoes A, Nichol KL, Hak E. Quantifying the potential role of unmeasured confounders: The example of influenza vaccination. *Int J Epidemiol* 2008; 37(6): 1422-9.
19. Schneeweiss S. Sensitivity analysis and external adjustment for unmeasured confounders in epidemiologic database studies of therapeutics. *Pharmacoepidemiol Drug Saf* 2006; 15(5): 291-303.
20. 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; 17(3): 268-75.
21. Tannen RL, Weiner MG, Xie D. Replicated studies of two randomized trials of angiotensin-converting enzyme inhibitors: further empiric validation of the 'prior event rate ratio' to adjust for unmeasured confounding by indication. *Pharmacoepidemiol Drug Saf* 2008; 17(7): 671-85.
22. Schneeweiss S, Stürmer T, Maclure M. Case-crossover and case-time-control designs as alternatives in pharmacoepidemiologic research. *Pharmacoepidemiol Drug Saf* 1997; 6(SUPPL. 3): S51-9.
23. Maclure M, Fireman B, Nelson JC, Hua W, Shoaibi A, Paredes A, Madigan D. When should case-only designs be used for safety monitoring of medical products? *Pharmacoepidemiol Drug Saf* 2012; 21(SUPPL. 1): 50-61.
24. Farrington CP. Relative incidence estimation from case series for vaccine safety evaluation. *Biometrics* 1995; 51(1): 228-35.
25. Whitaker HJ, Farrington CP, Spiessens B, Musonda P. Tutorial in biostatistics: The self-controlled case series method. *Stat Med* 2006; 25(10): 1768-97.
26. Suissa S. The case-time-control design. *Epidemiology* 1995; 6(3): 248-53.
27. Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative Controls: A tool for detecting confounding and bias in observational studies. *Epidemiology* 2010; 21(3): 383-8.
28. Flanders WD, Klein M, Darrow LA, et al. A method for detection of residual confounding in time-series and other observational studies. *Epidemiology* 2011; 22(1): 59-67.
29. Lee W-. Detecting and correcting the bias of unmeasured factors using perturbation analysis: A data-mining approach. *BMC Medical Research Methodology* 2014; 14:18(1).
30. Greenland S. An introduction to instrumental variables for epidemiologists. *Int J Epidemiol* 2000; 29(4): 722-9.
31. Hernán MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? *Epidemiology* 2006; 17(4): 360-72.

32. Martens EP, Pestman WR, De Boer A, Belitser SV, Klungel OH. Instrumental variables: Application and limitations. *Epidemiology* 2006; 17(3): 260-7.
33. Davies NM, Smith GD, Windmeijer F, Martin RM. Issues in the reporting and conduct of instrumental variable studies: A systematic review. *Epidemiology* 2013; 24(3): 363-9.
34. Chen Y, Briesacher BA. Use of instrumental variable in prescription drug research with observational data: a systematic review. *J Clin Epidemiol* 2011; 64(6): 687-700.
35. Nordmann S, Biard L, Ravaud P, Esposito-Farèse M, Tubach F. Case-only designs in pharmacoepidemiology: a systematic review. *PloS one* 2012; 7(11): e49444.
36. Rassen JA, Brookhart MA, Glynn RJ, Mittleman MA, Schneeweiss S. Instrumental variables I: instrumental variables exploit natural variation in nonexperimental data to estimate causal relationships. *J Clin Epidemiol* 2009; 62(12): 1226-32.
37. 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. *Can Med Assoc J* 2007; 176(5): 627.
38. Swanson SA, Hernán MA. Commentary: How to report instrumental variable analyses (suggestions welcome). *Epidemiology* 2013; 24(3): 370-4.
39. Groenwold RHH, Hak E, Klungel OH, Hoes AW. Instrumental Variables in Influenza Vaccination Studies: Mission Impossible?! *Value Health* 2010; 13(1): 132-7.
40. VanderWeele TJ, Tchetgen Tchetgen EJ, Cornelis M, Kraft P. Methodological challenges in mendelian randomization. *Epidemiology* 2014; 25(3): 427-35.
41. Ionescu-Ittu R, Abrahamowicz M, Pilote L. Treatment effect estimates varied depending on the definition of the provider prescribing preference-based instrumental variables. *J Clin Epidemiol* 2012; 65(2): 155-62.
42. Ionescu-Ittu R, Delaney JAC, Abrahamowicz M. Bias–variance trade-off in pharmacoepidemiological studies using physician-preference-based instrumental variables: a simulation study. *Pharmacoepidemiol Drug Saf* 2009; 18(7): 562-71.
43. Cai B, Small DS, Have TRT. Two-stage instrumental variable methods for estimating the causal odds ratio: Analysis of bias. *Stat Med* 2011; 30(15): 1809-24.
44. Davies NM, Gunnell D, Thomas KH, Metcalfe C, Windmeijer F, Martin RM. Physicians' prescribing preferences were a potential instrument for patients' actual prescriptions of antidepressants. *J Clin Epidemiol* 2013; 66(12): 1386-96.
45. Brookhart MA, Rassen JA, Schneeweiss S. Instrumental variable methods in comparative safety and effectiveness research. *Pharmacoepidemiol Drug Saf* 2010; 19(6): 537-44.
46. Brookhart MA, Schneeweiss S. Preference-based instrumental variable methods for the estimation of treatment effects: Assessing validity and interpreting results. *Int J Biostat* 2007; 3(1): 14.
47. Rassen JA, Brookhart MA, Glynn RJ, Mittleman MA, Schneeweiss S. Instrumental variables II: instrumental variable application-in 25 variations, the physician prescribing preference generally was strong and reduced covariate imbalance. *J Clin Epidemiol* 2009; 62(12): 1233-41.
48. Rassen JA, Schneeweiss S, Glynn RJ, Mittleman MA, Brookhart MA. Instrumental variable analysis for estimation of treatment effects with dichotomous outcomes. *Am J Epidemiol* 2009; 169(3): 273-84.
49. Palmer TM, Sterne JAC, Harbord RM, et al. Instrumental variable estimation of causal risk ratios and causal odds ratios in mendelian randomization analyses. *Am J Epidemiol* 2011; 173(12): 1392-403.
50. Burgess S. Identifying the odds ratio estimated by a two-stage instrumental variable analysis with a logistic regression model. *Stat Med* 2013; 32(27): 4726-47.
51. 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; 54(11): 3390-8.

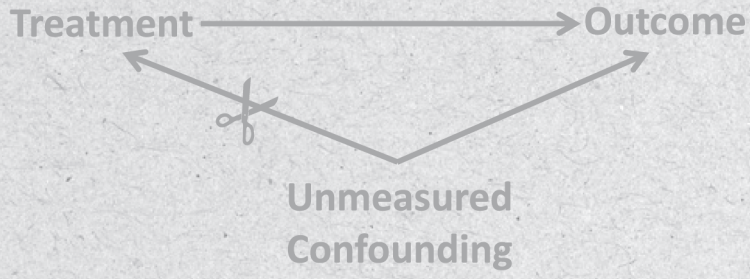
52. Pratt N, Roughead EE, Ryan P, Salter A. Antipsychotics and the risk of death in the elderly: An instrumental variable analysis using two preference based instruments. *Pharmacoepidemiol Drug Saf* 2010; 19(7): 699-707.
53. Huybrechts KF, Gerhard T, Franklin JM, Levin R, Crystal S, Schneeweiss S. Instrumental variable applications using nursing home prescribing preferences in comparative effectiveness research. *Pharmacoepidemiol Drug Saf* 2014; 23(8): 830-8. doi:10.1002/pds.3611.
54. Davies NM, Smith GD, Windmeijer F, Martin RM. COX-2 selective nonsteroidal anti-inflammatory drugs and risk of gastrointestinal tract complications and myocardial infarction: An instrumental variable analysis. *Epidemiology* 2013; 24(3): 352-62.
55. Stock JH, Wright JH, Yogo M. A survey of weak instruments and weak identification in generalized method of moments. *J Bus Econ Stat* 2002; 20(4): 518-29.
56. Angrist JD, Pischke JS. Mostly harmless econometrics: An empiricist's companion. Princeton Univ Pr; 2008.
57. Staiger D, Stock JH. Instrumental variables regression with weak instruments. *Econometrica* 1997; 65(3): 557-86.
58. Bound J, Jaeger DA, Baker RM. Problems with instrumental variables estimation when the correlation between the instruments and the endogeneous explanatory variable is weak. *J Am Stat Assoc* 1995; 90(430): 443-50.
59. Abbing-Karahagopian V, Huerta C, Souverein P, et al. Antidepressant prescribing in five European countries: application of common definitions to assess the prevalence, clinical observations, and methodological implications. *Eur J Clin Pharmacol* 2014; 70(7): 849-57.
60. Trifirò G, Coloma PM, Rijnbeek PR, et al. Combining multiple healthcare databases for postmarketing drug and vaccine safety surveillance: Why and how? *J Intern Med* 2014; 275(6): 551-61.
61. García Rodríguez LA, Herings R, Johansson S. Use of multiple international healthcare databases for the detection of rare drug-associated outcomes: A pharmacoepidemiological programme comparing rosuvastatin with other marketed statins. *Pharmacoepidemiol Drug Saf* 2010; 19(12): 1218-24.
62. Weiner MG, Xie D, Tannen RL. Replication of the Scandinavian Simvastatin Survival Study using a primary care medical record database prompted exploration of a new method to address unmeasured confounding. *Pharmacoepidemiol Drug Saf* 2008; 17(7): 661-70.
63. Tannen RL, Weiner MG, Xie D. Use of primary care electronic medical record database in drug efficacy research on cardiovascular outcomes: Comparison of database and randomised controlled trial findings. *BMJ (Online)* 2009; 338(7691): 395-9.
64. Yu M, Xie D, Wang X, Weiner MG, Tannen RL. Prior event rate ratio adjustment: Numerical studies of a statistical method to address unrecognized confounding in observational studies. *Pharmacoepidemiol Drug Saf* 2012; 21(SUPPL.2): 60-8.
65. Whitaker HJ, Hocine MN, Farrington CP. The methodology of self-controlled case series studies. *Stat Methods Med Res* 2009; 18(1): 7-26. doi:10.1177/0962280208092342 [doi].
66. Hua W, Sun G, Dodd CN, et al. A simulation study to compare three self-controlled case series approaches: Correction for violation of assumption and evaluation of bias. *Pharmacoepidemiol Drug Saf* 2013; 22(8): 819-25.
67. Paddy Farrington C, Anaya-Izquierdo K, Whitaker HJ, Hocine MN, Douglas I, Smeeth L. Self-controlled case series analysis with event-dependent observation periods. *Journal of the American Statistical Association* 2011; 106(494): 417-26.

68. Weldeselassie YG, Whitaker HJ, Farrington CP. Use of the self-controlled case-series method in vaccine safety studies: Review and recommendations for best practice. *Epidemiol Infect* 2011; 139(12): 1805-17.
69. Farrington CP, Whitaker HJ, Hocine MN. Case series analysis for censored, perturbed, or curtailed post-event exposures. *Biostatistics* 2009; 10(1): 3-16.

Unmeasured confounding IV PERR SCCS Observational studies
PROTECT Causal Inference Statistical methods পরিসংখ্যানগত পদ্ধতি
Statistische methoden IV PERR SCCS Pseudo Randomization
Pharmacoepidemiology IV Analyses PERR SCCS
General Practice Databases Residual Confounding
Electronic Healthcare Databases Epidemiology
Time-fixed Confounding Time-fixed Exposure
Time-varying Exposure
Time-varying Confounding
Non-randomized Studies
IV PERR SCCS
PROTECT 2SLS
IV IV

CHAPTER II

INTRODUCTION OF INSTRUMENTAL VARIABLE ANALYSIS



CHAPTER 2.1

Instrumental Variable Analysis in Randomized Trials with Non-compliance and Observational Pharmacoepidemiological Studies

Rolf HH Groenwold, Md Jamal Uddin, Kit CB Roes, Anthonius de Boer, Elena Rivero-Ferrer, Elisa Martin, Nicolle M Gatto, Olaf H Klungel

OA Epidemiology 2014 May 09; 2(1):9.

ABSTRACT

BACKGROUND: Instrumental variable (IV) analysis potentially accounts for unmeasured confounding in observational studies, but it can also control for non-compliance in randomized trials. We aimed to describe the conceptual framework of the IV analysis with its limitations in pharmacoepidemiology.

METHODS/ RESULTS: IV analysis requires that the IV is related to treatment status, yet independent of confounders of the treatment-outcome relation. This implies that in pharmacoepidemiological scenarios where IV analysis is needed the most (because of strong unmeasured confounding), IVs will typically be weakly associated with treatment. Furthermore, IV analysis assumes that the IV affects the outcome only through the treatment under study. A common IV in pharmacoepidemiological studies is the physician prescribing preference, which for the latter assumption implies that physicians only differ in their preference for the treatment under study, but they do not differ with respect to e.g., preferences for concomitant treatments, skills, organization of their practice, etc. Assumptions underlying IV assumptions need thorough evaluation before proceeding with IV analyses. Here, IV analysis is illustrated, its key assumptions explained and the utility of IVs for observational pharmacoepidemiological studies is discussed.

CONCLUSIONS: The validity and applicability of IV analysis in observational pharmacoepidemiological studies still have to be established, which requires more applications of IV analysis and debate on the likelihood of the assumptions underlying IV analysis.

INTRODUCTION

Observational studies of the effects of medical interventions (e.g., pharmacological treatment) are prone to confounding. Different methods are available to control for confounding, including restriction, matching, multivariable regression analysis, propensity score analysis, and inverse probability weighting.[1] What these methods have in common, is that they can control for measured confounders, but not for unmeasured confounders. Instrumental variable (IV) analysis, on the other hand, has been proposed as a method to control for unmeasured confounding in observational studies. In this review, we will discuss the limitations of commonly used IVs to control for unmeasured confounding in pharmacoepidemiology. First, IV analysis is described conceptually and illustrated by non-compliance in a randomized trial.

Causal diagrams of observational studies and randomized trials

Figure 1 shows several directed acyclic graphs (DAGs), also referred to as causal diagrams. For a detailed explanation of DAGs, we refer to the literature, e.g.,[2,3] Here, it suffices that causal relations between variables are represented by directed arrows from cause to effect

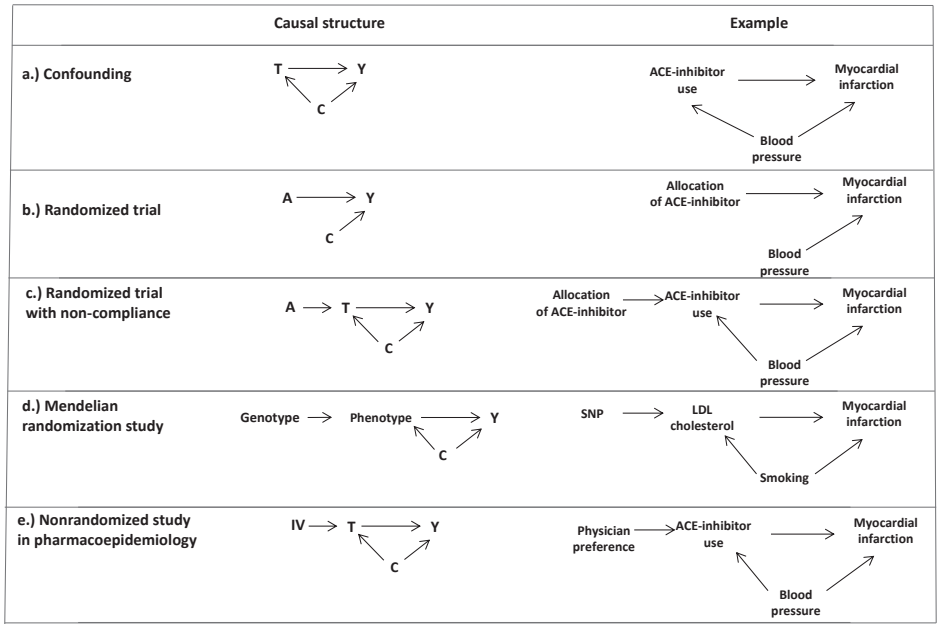


Figure 1. Directed acyclic graphs of randomized and observational studies

Abbreviations: T: treatment; Y: outcome; C: confounders; A: random treatment assignment, IV: instrumental variable. SNP: single-nucleotide polymorphism.

and all causal relations of the treatment and outcome are represented. The DAG in Figure 1a shows the typical structure of confounding. The allocation of treatment (T, e.g., treatment with an ACE-inhibitor) and outcome (Y, e.g., myocardial infarction) share a common cause (C, e.g., pre-treatment blood pressure). There is a so-called 'back-door path' from treatment to outcome, via the confounder C. Ignoring this back-door path when estimating the relation between treatment and outcome may result in a bias (i.e., confounding).

However, if one of the arrows from the confounder to either treatment or outcome is absent, there is no back-door path and hence no confounding. This is depicted in Figure 1b, which represents an ideal randomized controlled trial. Because treatment allocation is random, it is independent of subject characteristics and hence there is no arrow from C to T.

In reality, in a randomized trial, adherence to the randomly allocated treatment may not be perfect. Hence, treatment allocation (A) and actual treatment status (T) may not be identical (Figure 1c). Note that treatment allocation is still a random process (hence independent of C), yet treatment use need not be a random process. The latter is reflected by the arrow between C and T. An analysis of actual treatment may therefore be biased (due to confounding by C), yet an analysis of treatment allocation (i.e., intention-to-treat analysis) will on average be unbiased.

Non-compliance in a randomized trial

The intention-to-treat (ITT) analysis of a randomized trial provides an unbiased estimate of the effect of treatment allocation, rather than the effect of actual treatment use. If the treatment is effective, the ITT analysis underestimates the effect of treatment use, when there is considerable non-compliance.[4] However, by taking the extent of non-compliance into account, one can estimate what the treatment effect would be under perfect compliance. We illustrate this using numerical examples.

Table 1 shows three numerical examples of randomized trials. In the first scenario, there is perfect compliance: all subjects allocated to the experimental treatment actually receive the experimental treatment and all allocated to the control treatment receive the control treatment. Hence, the estimate of the effect of treatment allocation equals the effect estimate of actual treatment received: risk difference (RD) = -0.25.

Table 1. Numerical example of trials with no or partial non-compliance

Scenario 1: no non-compliance				
Treatment assigned		Treatment received	No. subjects	No. events
T = 0		T = 0	1000	500
T = 1		T = 1	1000	250
Scenario 2: partial non-compliance, unrelated to any cause of the outcome				
Treatment assigned		Treatment received	No. subjects	No. events
T = 0		T = 0	600	300
		T = 1	400	100
T = 1		T = 0	200	100
		T = 1	800	200
Scenario 3: partial non-compliance, related to blood pressure at baseline				
Treatment assigned	Blood pressure	Treatment received	No. subjects	No. events
T = 0	High	T = 0	300	225
		T = 1	300	150
	Low	T = 0	320	160
		T = 1	80	20
T = 1	High	T = 0	120	90
		T = 1	480	240
	Low	T = 0	200	100
		T = 1	200	50

Abbreviations: T = 0: control treatment; T = 1: experimental treatment.

The second scenario is that of a randomized trial with non-compliance: 60% and 80% of those assigned the control treatment and the experimental treatment, respectively, comply with the assigned treatment. The ITT effect can be estimated as RD = 300/1000 – 400/1000 = -0.1, which underestimates the effect that would be observed under perfect compliance (scenario 1). To obtain the effect that would be observed under perfect compliance, the ITT effect needs to be extrapolated to a situation with full compliance. This can be achieved by dividing the ITT effect by the difference in the observed probabilities of receiving treatment between the two treatment allocation groups: -0.1 / (800/1000 – 400/1000) = -0.1 / 0.4 = -0.25, which indeed equals the effect that is observed under perfect compliance.[5,6]

A graphical representation of this procedure is given in Figure 2. The observed risks among the two treatment allocation groups (0.4 for the control group and 0.3 for the experimental group) are plotted against the probabilities of actually receiving treatment among those two groups (0.4 and 0.8 for the control and experimental treatment groups, respectively). These two points are then connected. The risk difference that would be observed under perfect compliance can be obtained by extrapolating this line to the point at which the

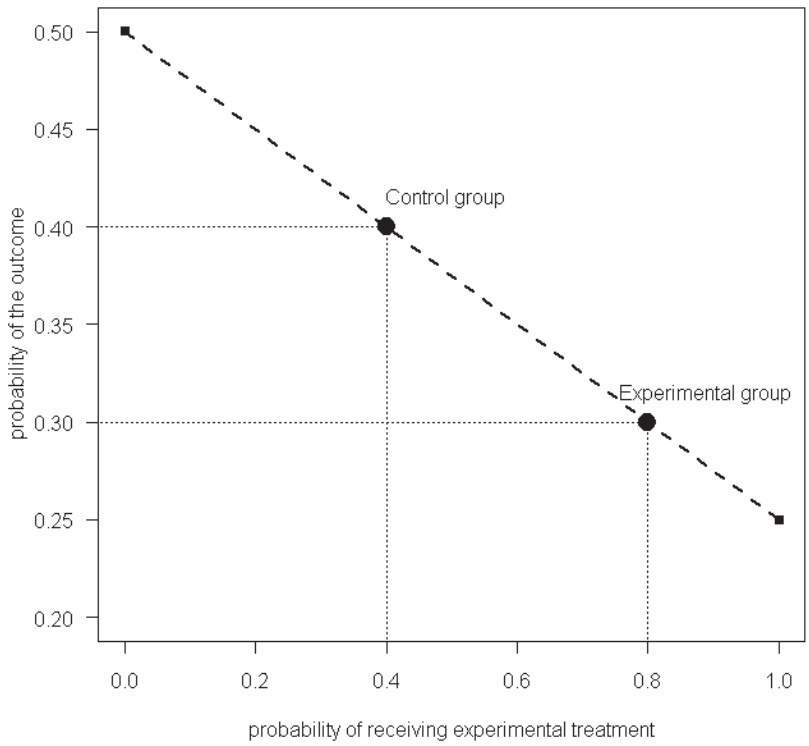


Figure 2. Graphical representation of IV analysis of a randomized trial with non-compliance

probability of receiving treatment is either 0 or 1. The difference between those two extremes can be read off the y-axis and is the risk difference that would be observed under perfect compliance.

In scenario 2, compliance differs between treatment arms, but within treatment arms it is a random process. However, the method to account for non-compliance that was described also works if actual treatment status depends on random treatment allocation as well as risk factors for the outcome (the DAG in Figure 1c), which is illustrated by scenario 3 (Table 1). The actual treatment received now depends on treatment allocation, but also on blood pressure: regardless of treatment allocation, those with a high pre-treatment blood pressure are more likely to use treatment compared to those with a low blood pressure. Note that the distribution of blood pressure is the same in the two randomization groups. The ITT effect (RD = 480/1000 – 555/1000 = -0.075) again underestimates the treatment effect that would be observed under perfect compliance. However, the ITT effect can be adjusted by the difference in the observed probabilities of receiving experimental treatment between the two treatment allocation groups to obtain the treatment effect under perfect compliance: -0.075 / (680/1000 – 380/1000) = -0.075 / 0.3 = -0.25.

In scenario 3, an analysis that is stratified by blood pressure will also yield an unbiased estimate of the treatment effect under perfect compliance. However, this obviously requires that blood pressure is actually measured, whereas the analysis outlined in scenario 2 above can also be conducted when blood pressure is unmeasured; hence in the presence of unmeasured confounding.

Assumptions of instrumental variable analysis

The procedure outlined above to account for non-compliance in randomized trials (in scenario 2 above) is a particular form of IV analysis. IV analysis can account for non-compliance in a randomized trial to the extent that there is some contrast in the probability of treatment use between the two randomization groups.[7] This is the first main assumption of IV analysis, which can be summarized as 1.) an IV predicts treatment status. Several statistical measures are available to quantify the relation between IV and treatment, including correlation, odds ratio, and proportion of explained variance.[8,9] The importance of this assumption can be easily understood by looking at Figure 2. If the two points are very close to each other, extrapolating the line between the two points will become a very inaccurate process. The further away these points are, the more precise the extrapolation will be. A weak association between IV and treatment becomes less influential in larger samples. Although the two point are close to each other, they have a large precision (due to the large sample size) which will attenuate the instability of the extrapolation. The relation between IV and exposure is reflected by the arrow between these variables in the DAGs in Figure 1.

There are two other main assumptions underlying IV analysis: 2.) an IV affects the outcome only through the treatment and 3.) an IV is independent of confounders of the treatment-outcome relation. For the DAGs in figures 1b and 1c, this implies that the observed effect

of treatment assignment on the outcome runs completely through the indicated arrows, i.e., there are no unrepresented associations (arrows) between the IV and the outcome (assumption 2), nor are there any back-door paths from the IV to treatment status or from the IV to the outcome (assumption 3). In a perfect randomized trial, blinding is used in an attempt to meet assumption 2 (which ensures that treatment arms will remain comparable during follow-up), whereas randomization is used to the meet assumption 3.

A causal effect can still be estimated if the assumption of 'no relation between IV and confounders' (assumption 3) can be relaxed. There should be no unmeasured confounders of relation between IV and exposure and the relation between IV and outcome. In the numerical examples above (scenario 2), the exposure effect was estimated as the ratio of the ITT effect and the relation between IV status and exposure status. If either risk difference is biased (by unmeasured confounding), their ratio may be biased as well. However, in the absence of other biases, if both elements of the ratio are adjusted for measured confounders of those risk differences, the ratio may yield an unbiased estimate of the exposure effect. The assumption of no unmeasured confounders cannot be proved, but it may be falsified in the data.[10] An observed imbalance in measured confounders within IV strata may suggest that unmeasured confounders are imbalanced as well, thus invalidating IV analysis. Importantly, the bias due to unmeasured confounding can be much larger when conducting IV analysis compared to conventional analysis.[10]

The ratio method outlined above is just one of many possible statistical approaches to IV analysis. Furthermore, additional assumptions are required to interpret the estimated IV effect as causal, for example, assumptions related to homogeneity of the treatment effect. These additional assumptions as well as more flexible IV analytical methods are beyond the scope of this review and we refer to the literature for more details.[11-17]

Instrumental variables in observational studies

The application of IV analysis can be extended beyond non-compliance in randomized trials. In fact, randomized trials are just one of many possible fields of application. In observational studies, a variable that is not randomly allocated by the investigator, yet fulfils the assumptions of an IV, can act in the same way as random treatment allocation in a randomized trial.

For example, in a study of the relation between HDL-cholesterol levels (exposure) and myocardial infarction (outcome), a genetic polymorphism that increases HDL-cholesterol levels (and does not affect LDL-cholesterol or other cardiovascular risk factors) was used as IV.[18] Genetic polymorphisms are randomly distributed in populations and are in that respect similar to random treatment allocation in a randomized trial (Figure 1d). Studies that make use of this phenomenon are called Mendelian randomization studies.[18-20] In contrast to randomized trials, however, the relation between the IV and exposure is typically weak in Mendelian randomization studies, therefore requiring (very) large sample sizes,[21,22] as explained above.

Full knowledge of the biological mechanism by which the genetic polymorphism acts (e.g., does the polymorphism only affect HDL-cholesterol levels or also other biomarkers which may affect the risk of the myocardial infarction) is necessary to be confident that the assumptions of IV analysis hold.[23]

Instrumental variables in pharmacoepidemiology

Pharmacoepidemiological studies are often conducted in large databases of electronic healthcare records, which provide detailed information about for example co-morbidity and co-medication, but often have limited information about health behaviour (e.g., smoking, exercise, and dietary habits).The latter leads to a potential for unmeasured confounding, which may be overcome by IV analysis.

A review of IV analysis in pharmacoepidemiology, published in 2011, identified 5 types of instrumental variables that are typically used: regional variation, facility prescribing patterns, physician preference, patient history / financial status, calendar time.[24] Facility prescribing patterns together with physician preference are together the most commonly used IV in pharmacoepidemiology. In the remainder we focus on the IV physician prescribing preference (or physician preference).

Figure 1e shows the assumed causal structure of a study using physician preference as an IV. The IV physician preference can be defined in different ways (Figure 3). For example, in a study in which two drugs are compared against each other (A vs. B), all subjects treated with either A or B, from a number of participating practices, are enrolled in the study. For each physician the preference can then be defined as the number of prescriptions of drug A (n_A) compared to all prescriptions ($n_A + n_B$) made by that physician (third column in Figure 3). If the preference changes over time, one overall preference per physician may not be appropriate.[25,26] Instead, for each physician the percentage of prescriptions of drug A can be determined per year, or per quarter, to better account for possible changes over time. Ultimately, the prescription that was issued for the last patient before the current one could be used as a proxy for the preference of a physician at that moment in time.[25,27] If the last patient was prescribed drug A, then apparently the physician's preference at that moment is in favour of drug A (fourth column in Figure 3).

Interplay between IV assumptions

When applying IV analysis, researchers must demonstrate, or explicitly argue why, the assumptions of IV analysis hold. It is straightforward to check whether physician preference is indeed related to actual treatment. For example, IV status should predict to a considerable extent the actual treatment status. It is hard to provide cut-points that universally apply, but simulations suggest that for example the odds ratio between a binary IV and a binary treatment in a typical pharmacoepidemiological study should exceed 2 (note that this value depends on sample size, but not on statistical significance).[9] The assumption of independence between IV and confounders can be checked at least for the measured confounders, by making a comparison of confounders between levels of physician preference.[10]

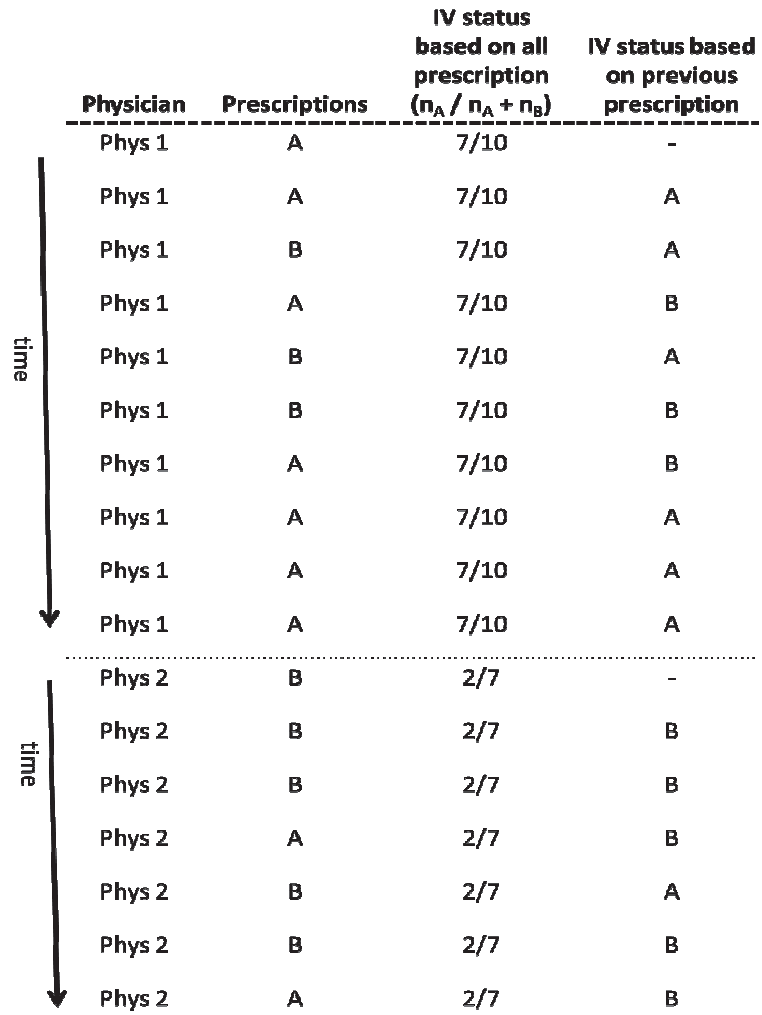


Figure 3. Definitions of common ways of building the instrumental variable physician's prescribing preference

According to the DAGs in Figure 1e, both the IV and potential confounders of the exposure-outcome relation affect actual treatment status. This means that if the proportion of explained variation in the treatment due to the IV is relatively large, there is little variation in treatment left that can be attributed to the confounders.[28] And vice versa, if the proportion of explained variation in the treatment due to confounders is relatively large, there is little variation in exposure left that can be attributed to the IV. Hence, in case of strong confounding, any IV that is independent of the confounders will only be weakly related to treatment. Only if the amount of confounding is limited, one may identify a strong IV. An

exception may be a situation in which the confounder-treatment association is relatively weak, yet the amount of confounding is substantial due to a very strong confounder-outcome association. Thus, particularly in those situations where IV analysis is needed the most to deal with (strong) unmeasured confounding, IVs will typically be weakly associated with treatment and thus require large sample sizes.

When treatment options are clearly spelled-out in clinical guidelines, any variation in prescribing rates between physicians will likely be small. In those situations in which the preference of a physician can lead to large variations in prescribing behaviour, apparently guidelines aren't that strict, which may be because there is clinical equipoise; i.e., there are no apparent risks or benefits related to one drug compared to the other. Consequently, treatment will probably not be prescribed very selectively, which means that the potential for confounding will be small.

The IV assumption that physician preference affects the outcome only through the treatment cannot be checked in the data. This assumption implies that physicians only differ in their preference for the treatment under study, but they do not differ with respect to all kinds of other aspects (e.g. preferences for concomitant treatments, skills, organization of their practice, etc) that may affect the outcome.[12] If they only differ in that respect, however, differences in preferences will likely be small and hence the IV will be weakly related to exposure status. On the other hand, in case of really distinct preferences (strong IV), physicians will likely differ in more respects than only their preference for that particular treatment, which impairs the validity of the IV physician preference. Obviously, physicians can be different in terms of e.g. sex and age. But as long as the sex and age of a physician are not related to the outcome, such differences will not affect the validity of the IV.

CONCLUSIONS

The use of instrumental variable analysis is clearly indicated in trials with non-compliance and in Mendelian randomization studies. However, its validity and applicability in observational studies of the effects of (pharmacological) treatments still have to be established. This requires more applied studies using different types of possible IVs. For each of these, the assumptions underlying IV analysis have to be thoroughly assessed and those assumptions that cannot be verified using the data have to be debated.[29] Importantly, physician preference is not the only possible IV for pharmacoepidemiologic studies. Differential implementation of guidelines between (similar!) regions, or evaluating the implementation of guidelines (before-after comparison) may provide valid IVs. We propose that new IVs are considered that allow for estimating unbiased treatment effects of safety and effectiveness in pharmacoepidemiology.

REFERENCES

1. Klungel OH, Martens EP, Psaty BM, Grobbee DE, Sullivan SD, Stricker BH, Leufkens HG, de Boer A. Methods to assess intended effects of drug treatment in observational studies are reviewed. *J Clin Epidemiol*. 2004;57(12):1223-31.
2. Rothman KJ, Greenland S, Lash TL. Modern Epidemiology. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2008.
3. Pearl J. Causal diagrams for empirical research. *Biometrika* 1995;82(4):669-688.
4. Hernan MA, Hernandez-Dıaz S. Beyond the intention-to-treat in comparative effectiveness research. *Clin Trials*. 2012;9(1):48-55.
5. Greenland S. An introduction to instrumental variables for epidemiologists. *Int J Epidemiol*. 2000;29(4):722-9.
6. Sexton M, Hebel JR. A clinical trial of change in maternal smoking and its effect on birth weight. *JAMA*. 1984;251(7):911-5.
7. Efron B, Feldman D. Compliance as an explanatory variable in clinical trials. *J Am Stat Assoc*. 1991; 86(413):9-26.
8. Bound J, Jaeger DA, Baker RM. Problems with instrumental variables estimation when the correlation between the instruments and the endogeneous explanatory variable is weak. *J Am Stat Assoc*. 1995;90(430):443-50.
9. Uddin MJ, Groenwold RH, de Boer A, Belitser SV, Roes KC, Hoes AW, Klungel OH. Performance of instrumental variable methods in cohort and nested case-control studies: a simulation study. *Pharmacoepidemiol Drug Saf*. 2014;23(2):165-77.
10. Ali MS, Uddin MJ, Groenwold RHH, Pestman WR, Belitser SV, Hoes AW, de Boer A, Roes KCB, Klungel OH. Quantitative falsification of instrumental variables assumption using balance measures. *Epidemiology*, 2014; 25(5):770-72.
11. Hernán MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? *Epidemiology*. 2006;17(4):360-72.
12. Baiocchi M, Cheng J, Small DS. Instrumental variable methods for causal inference. *Stat Med*. 2014;33(13):2297-340.
13. Palmer TM, Sterne JAC, Harbord RM, et al. Instrumental variable estimation of causal risk ratios and causal odds ratios in mendelian randomization analyses. *Am J Epidemiol*. 2011;173(12):1392-403.
14. Clarke PS, Windmeijer F. Identification of causal effects on binary outcomes using structural mean models. *Biostatistics*. 2010;11(4):756-70.
15. Angrist J, Imbens G, Rubin DB. Identification of causal effects using instrumental variables. *J Am Stat Assoc*. 1996;91(434):444-72.
16. Robins J, Rotnitzky A. Estimation of treatment effects in randomised trials with non-compliance and a dichotomous outcome using structural mean models. *Biometrika*. 2004;91(4):763-83.
17. Voight BF, Peloso GM, Orho-Melander M, Frikke-Schmidt R, Barbalic M, Jensen MK, et al. Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. *Lancet*. 2012;380(9841):572-80.
18. Burgess S, Butterworth A, Malarstig A, Thompson SG. Use of Mendelian randomisation to assess potential benefit of clinical intervention. *BMJ*. 2012;345:e7325.
19. Didelez V, Sheehan N. Mendelian randomization as an instrumental variable approach to causal inference. *Stat Methods Med Res*. 2007;16(4):309-30.
20. Lawlor DA, Harbord RM, Sterne JA, Timpson N, Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat Med*. 2008;27(8):1133-63.

21. Freeman G, Cowling BJ, Schooling CM. Power and sample size calculations for Mendelian randomization studies using one genetic instrument. *Int J Epidemiol*. 2013;42(4):1157-63.
22. Burgess S. Sample size and power calculations in Mendelian randomization with a single instrumental variable and a binary outcome. *Int J Epidemiol*. 2014 Mar 6
23. Vanderweele TJ, Tchetgen Tchetgen EJ, Cornelis M, Kraft P. Methodological challenges in mendelian randomization. *Epidemiology*. 2014;25(3):427-35.
24. Chen Y, Briesacher BA. Use of instrumental variable in prescription drug research with observational data: a systematic review. *J Clin Epidemiol*. 2011;64(6):687-700.
25. 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;17(3):268-75.
26. Abrahamowicz M, Beauchamp M, Ionescu-Ittu R, Delaney JAC. Pilote L. Reducing the variance of the prescribing preference-based instrumental variable estimates of the treatment effect. *Am J Epidemiol*. 2011;174(4):494-502.
27. Rassen JA, Brookhart MA, Glynn RJ, Mittleman MA, Schneeweiss S. Instrumental variables II: instrumental variable application-in 25 variations, the physician prescribing preference generally was strong and reduced covariate imbalance. *J Clin Epidemiol*. 2009;62(12):1233-41.
28. Martens EP, Pestman WR, de Boer A, Belitser SV, Klungel OH. Instrumental variables: application and limitations. *Epidemiology*. 2006;17(3):260-7.
29. Swanson SA, Hernán MA. Commentary: How to report instrumental variable analyses (suggestions welcome). *Epidemiology*. 2013; 24(3): 370-374.

CHAPTER 2.2

Instrumental Variable Analysis: A Methodological Review for Epidemiologists

Md Jamal Uddin, Rolf HH Groenwold, Anthonius de Boer, Svetlana V Belitser, Kit CB Roes, Olaf H Klungel

Submitted, Journal of Clinical Epidemiology

ABSTRACT

BACKGROUND: Instrumental variables (IV) analysis has been used to control for unmeasured confounding in observational studies. To perform a proper IV analysis, it is important to understand the methodology underlying IV analysis. We aim to provide an overview of IV methods and to provide guidance on when different methods may be applicable together with some advantages/disadvantages of the methods.

METHODS: We reviewed methods for IV analysis for continuous as well as binary outcome, exposure, and IV.

RESULTS: Two-stage least squares is the method of first choice if exposure and outcome are both continuous and show a linear relation. In case of a nonlinear relation, two-stage residual inclusion is a suitable alternative. In settings with binary outcomes as well as nonlinear relations between exposure and outcome, generalized method of moments (GMM), structural mean models (SMM), and bivariate probit models perform well, yet GMM and SMM are generally more robust. The standard errors of the IV estimate can be estimated using a robust or bootstrap method. For all methods, the assumptions underlying IV analysis are crucial and should be justified based on available statistical tools and clinical knowledge.

CONCLUSIONS: Researchers should be aware of the assumptions underlying IV analysis when interpreting IV estimates and chose the appropriate IV method, depending on the type of outcome, exposure, and instrumental variables.

INTRODUCTION

Instrumental variables (IVs) analysis has been used to control for unmeasured confounding in observational (pharmaco-)epidemiological studies.[1–6] IV analysis have primarily been used in economics and social science research, as a tool for causal inference, but have begun to appear in epidemiologic research over the last decade.[2,7–11] In economic research, outcomes are often continuous in nature and therefore linear IV methods are well suited and widely used.[4,12–18] A review of the application of IV analysis in epidemiology indicated physician prescribing preference (PPP) as the most commonly used IV in this field.[9] This IV can be either continuous or dichotomous. In addition, epidemiologic research often involves binary exposures and outcomes.[16] Application of IV analysis in epidemiology may therefore require other analytical methods than those commonly used in economic research. Hence, it is important to understand the methodology underlying IV analysis.

Our aim was to provide an overview of IV methods and to indicate which methods are appropriate in a given setting (i.e., given a certain combination of continuous and/or binary outcome, exposure and IV). In this review article, we synthesize current knowledge of IV methods for different types of IVs, exposures, and outcomes. After a general introduction to the assumptions underlying IV analysis, we will consecutively consider the different

combinations of binary and continuous IVs, exposures and outcomes and provide guidance on when different methods may be applicable together with some advantages/ disadvantages of the methods.

Instrumental variables

An IV is a variable that can be considered to mimic the treatment assignment process in a randomized study.[8,19–21] The IV is (strongly) related to exposure, and only related to the outcome through exposure. This resembles a RCT, in which treatment assignment typically almost perfectly coincides with the actual treatment received and (in case of a blinded trial) treatment assignment only affects the outcome through the received treatment (hence the term pseudo-randomisation that is used for IV methods). This implies that an IV is neither directly nor indirectly (e.g., through measured or unmeasured confounders) associated with the outcome.[6,9,12] Therefore, all measured and unmeasured confounders should on average be equally distributed among different levels of the IV (similar to an RCT). These assumptions are illustrated in Figure 1. Along with these basic assumptions, there are additional assumptions that are needed for point identification of IV estimators. These are the assumptions of a homogeneous treatment effects (i.e., treatment effects are constant across the study population), or monotonicity (the IV affects the treatment deterministically in one direction) in case of heterogeneous treatment effect or no effect modification by the IV.[22]

Notation and outline

Throughout this article, we use the following notation: Y denotes the outcome, X denotes exposure, and Z denotes the IV. C and U denote the (one or more) measured and unmeasured confounding variables, respectively. $\hat{X}$ denotes the predicted value of exposure. Finally, $\hat{\beta}_{IV}$ indicates the IV estimator, i.e., the estimator of the causal relation between exposure and outcome.

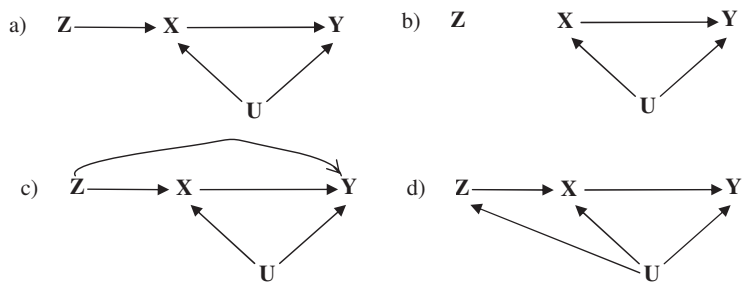


Figure 1. Schematic presentation of valid and invalid instrumental variables

X, Y, Z, and U denote the exposure, outcome, IV, and confounders, respectively. a) Z is associated with X and only related to Y through X (valid IV), b) Z is not associated with X (first IV assumption is violated), c) Z is not independent of Y given X and U, i.e. Z has a direct effect on Y (second IV assumption violated) d) Z is not independent of confounders, i.e. Z has an indirect effect on Y (third IV assumption is violated).

Different methods for IV analysis are summarized in this article, including the ratio estimator (RE); two-stage least squares (2SLS); three-stage least squares (3SLS); two-stage predictor substation (2SPS); two-stage residual inclusion (2SRI); linear probability model (LPM); two-stage logistic regression (2SLR); additive structural mean model (ASMM) and multiplicative structural mean model (MSMM); logistic structural mean model (LSMM); generalized method of moments with a linear model (LGMM) and a multiplicative model (MGMM); bivariate probit model (BPM); and other particular two-stage methods. A summary of the IV methods we discuss is provided in Table 1. We discuss each method once, but provide a table that shows all possible situations to apply the IV methods (Table 2). Although some methods may be applicable in other situations, we only present combinations that have been described commonly in the literature.

Continuous outcome, continuous exposure, binary IV

To assess the causal effect of maternal smoking (i.e., number of cigarettes smoked per day) on infant birth weight, an IV analysis was performed in a study in which an encouragement program to stop smoking was randomly allocated to pregnant women who smoked.[23] In this example, the IV was a binary variable (enrolment in the encouragement program, or no enrolment), whereas the exposure (number of cigarettes smoked per day) and the outcome (birth weight) were continuous.

In this setting, several methods such as RE, 2SLS, 3SLS, ASMMs, 2SPS, and 2SRI are possible to apply (Table 2). All these methods provide the same exposure effects in the case of a single IV, no other covariates, and linear association between X and Y. As the RE is commonly used and easy to apply with single IV, we describe only this method in this section.

Ratio estimator (RE)

In case of a continuous outcome, a continuous exposure and a binary IV, the RE (also called Wald[24] or grouping estimator) is expressed as:

$$\hat{\beta}_{IV} = \frac{\bar{y}_1 - \bar{y}_0}{\bar{x}_1 - \bar{x}_0} \qquad [1]$$

where $\bar{y}_1$ and $\bar{x}_1$ are the mean of *y* and *x*, respectively, when $z = 1$ and $\bar{y}_0$ and $\bar{x}_0$, when $z = 0$.

Continuous outcome, continuous exposure, continuous IV

In a study on the effect of childhood body mass index (BMI) on age of menarche, an genetic allele score was used as an IV. In this example, the IV (allele score), the exposure (childhood BMI) as well as the outcome (age of menarche) were continuous.[25]

In this setting, a particular case of ratio estimator can be used, which would then be defined as:

$$\hat{\beta}_{IV} = \frac{\hat{\beta}_{YZ}}{\hat{\beta}_{XZ}},$$

where $\hat{\beta}_{YZ}$ is the effect of Z on Y (obtained from e.g., a regression model) and $\hat{\beta}_{XZ}$ is the effect of Z on X (again from e.g., a regression model).[26] It is also possible to combine the

two estimations in a so-called two-stage IV method. Moreover, other methods e.g., 2SLS, 3SLS, 2SPS, 2SRI, and LGMM are also applicable here (Table 2) and with single IV and linear models, these methods provide similar exposure effects on outcome.[3,17,27–29] As 2SLS is mostly used in this setting, we describe this method here. Additionally, in order to account possible correlation between errors in the 2SLS, 3SLS is also suitable in this combination and hence explained below.

Two-stage least squares method (2SLS)

In two-stage methods, the results of a first-stage model are used in a second-stage model. The best known two-stage method for IV analysis is the 2SLS method which is traditionally used in IV analyses.[21,30,31] The 2SLS estimator can be obtained by the following models:

$$X_i = \alpha_0 + \alpha_z Z_i + \alpha_c C_i + \varepsilon_{1i}\text{; for } i = 1,2,.....n \quad [2]$$

$$\mathrm{Y_i} = \beta_0 + \beta_{\mathrm{IV}}\hat{\mathrm{X}}_\mathrm{i} + \beta_\mathrm{c}\mathrm{C_i} + \varepsilon_{2\mathrm{i}}\text{; for } i = 1,2,.....n \quad [3]$$

The first model estimates the effect of the IV on exposure, whereas in the second model outcomes are compared in terms of predicted exposure rather than the actual exposure. The latter model yields the estimated parameter, $\hat{\beta}_{IV}$, which is the IV estimator. In case of multiple IVs, information on these IVs can be simultaneously incorporated in model (2). Then, $\hat{\beta}_{IV}$ is the weighted average of the ratio estimators.[32] One of the conditions of this method is that the error term should be homoscedastic (homogeneity of variance). However, in case of heteroscedasticity, other methods (e.g., generalized method of moments, structural mean models) can be considered.[33] For multiple IVs, 2SLS provides biased estimates[32,34,35] and another method, e.g., limited information maximum likelihood (LIML),[36] can be an alternative.

Three-stage least squares method (3SLS)

The 3SLS was proposed by Zellner and Theil[37] and generalizes the 2SLS. Possible correlation of the errors (ε_1 and ε_2) in equations (2) and (3) is not taken into account by 2SLS. 3SLS accounts for the possible correlations between errors and may improve the efficiency of the estimator.[38,39] Unlike 2SLS, in which the coefficients of the two equations are estimated separately, in 3SLS all coefficients are estimated simultaneously. This requires three steps. The first-stage is similar to the 2SLS, i.e., a linear regression of X on Z to get $\hat{X}$. In the second-stage, the residuals of the second-stage 2SLS model are obtained to estimate the cross-model correlation matrix (correlation between error terms in both models). Finally, in the third-stage the estimated correlation matrix is used to obtain the IV estimator. When there is no correlation between the error terms of the 2SLS models, the 3SLS reduces to a 2SLS. However, 3SLS is more vulnerable to misspecification error since misspecification of one of the models in the first or second will affect the third stage model.[40]

Continuous outcome, binary exposure, binary IV

In observational studies of the effects of beta-blocker therapy vs. diuretics on blood pressure in hypertensive patients, the groups of patients that are registered with certain physicians

and receive beta-blocker vs. diuretics therapy may have similar characteristics, while the physicians may differ in their preference for either beta-blockers or diuretics. In that case, IV analysis may be possible, in which the outcome (blood pressure) is continuous, and the exposure (beta-blocker or diuretic therapy) as well as the IV (preference for beta-blockers or diuretics) are binary.[41]

Ratio estimator

When the Y is continuous and X and Z are binary, the ratio estimator can be defined as below:[42,43]

$$\hat{\beta}_{IV} = \frac{E[Y \mid Z = 1] - E[Y \mid Z = 0]}{p(X = 1 \mid Z = 1) - p(X = 1 \mid Z = 0)} \quad [4]$$

where $p(X = 1|Z = 1) - p(X = 1|Z = 0)$ is the difference in probability of being exposed for $Z = 1$ and $Z = 0$.

Structural mean models

Structural mean models (SMMs) explicitly use counterfactuals or potential outcomes,[28] which were originally proposed by Robins[44] in the context of RCTs with non-compliance to estimate the causal effects for the treated (exposed) individuals. SMMs are semi-parametric models and use IVs via G-estimation for identification and estimation of the causal parameter. This method involves the assumption of a conditional mean independence (CMI)[22,45–48] and does not make distributional assumptions about the exposure.[45] SMMs with an identity link is sometimes called additive SMMs and can be used for continuous outcomes (applicable in this setting) and multiplicative SMMs with log-linear model can be used for positive-valued/binary outcomes in order to estimate the causal risk ratio.[45,49] Additionally, the logistic structural mean model (LSMM) developed by Vansteelandt and Goetghebeur[50] and Robins and Rotnitzky[51] can also be used for binary outcome in order to estimate causal odds ratio.[45,49]

To handle continuous outcome data, the IV estimator from the additive SMMs can be expressed as equation (4) given that the assumptions of CMI and no effect modification by Z are fulfilled.[22,48,52,53] This estimator provides the average treatment effect (ATT) for the treated individuals.[18,45]

The advantage of this method is that it relaxes several of the modelling restrictions such as homogeneous treatment effects required by more classical methods such as RE/two-stage IV methods.[22,45] One of the key assumptions of this method is no effect modification, which is difficult to verify in practical situations.[53]

SMMs have been extended by Robins[46] to a general setting of structural nested mean models (SNMM) for repeated measures at multiple time points. The SMMs are a subclass of the SNMM.[44,54] When instruments, exposures, and confounders are time-dependent, SNMM can be used to estimate causal effects of exposure on the outcome.[22] Details and mathematical formulations of SMMs are described elsewhere.[22,45,49]

Apart from these two methods, 2SLS, 2SPS, and 2SRI are easy to apply. However, considering flexible modelling assumptions, additive SMM can be preferable in this combination.

Continuous outcome, binary exposure, continuous IV

Consider an observational study on the effects of adherence to beta-blocker therapy on blood pressure, in a setting with drug co-payment by patients.[55] In this situation, the outcome (blood pressure) is continuous, and exposure (adherent vs. non-adherent to beta-blocker therapy) is binary. The amount of drug co-payment by patients could serve as a (continuous) IV.

In this situation, 2SLS, 2SPS, and 2SRI are suitable to apply. As exposure is binary, the first-stage model can be a non-linear model. Hence, 2SPS and 2SRI are explained here though 2SRI is more preferable.

Two-stage predictor substitution

A two-stage method that can be considered in this setting is two-stage predictor substation, which is an extension of 2SLS to nonlinear models.[56–58] In the first-stage, a nonlinear least squares method (NLS) or any other consistent estimation technique is used to estimate the relation between the IV and exposure.[16] Then, the predicted exposure status from the first-stage model replaces the observed exposure as the principal covariate in the second-stage model on the outcome.[16,59] For a continuous exposure and outcome, 2SPS and 2SLS show similar results.[12,57]

Two- stage residual inclusion

Two-stage residual inclusion (2SRI) (also called control function estimator)[60] is another two-stage method and was first suggested by Hausman.[61] The general notion of the 2SRI is to include the error terms (residuals) from the first-stage model as an additional variable along with the exposure in the second-stage model.[62] The models in the first and second-stage can be either linear or nonlinear models. In case of linear models, the 2SRI estimate is equivalent to the 2SLS and 2SPS estimates.[17,59]

2SRI yields consistent estimates for both linear and nonlinear models.[29,63] The advantage of 2SRI over 2SLS is that 2SLS is only consistent when the second-stage model is linear, whereas this restriction does not hold for 2SRI.[16,64] Moreover, this method shows more precise estimates than 2SPS.[28]

Binary outcome, continuous exposure, binary IV

In a study of the effect of BMI (continuous exposure) on childhood asthma (dichotomous outcome), FTO genotype was considered as a (binary) IV.[45] BMI may be related to FTO genotype; the latter being randomly distributed in the population due to Mendel's laws and thus independent of measured and unmeasured confounders. Hence, this study possibly allows for IV analysis.

Table 2 shows that several methods are also applicable in this combination. Considering flexible modelling assumptions, we recommend to choose GMM with nonlinear models and MSMM. Since MSMM is explained before, we describe GMM in this section.

Generalized method of moments

When applying the generalized method of moments (GMM) a system of equations is set up, which is then solved numerically using computer algorithms. This technique was formalized by Hansen[65] and is a broad class of estimation methods that allow for a larger number of equations (moment conditions) than parameters[4,14,15] that are not possible in the MSMM and LSMM.[45] More clearly, the GMM allows for estimation of parameters in an over-identified model (number of IVs greater than the number of exposures). GMM with linear model can be similar to the ones used in 2SLS[66] but GMM is also a non-linear analogue of 2SLS[67] which is called multiplicative GMM. Detailed explanations can be found elsewhere.[4,14,45]

In general, the nonlinear optimum GMM estimator is asymptotically more efficient than 2SLS.[68] Since GMM is a moment based method without parametric assumptions, it is less prone to model misspecification than 2SLR or BPM when exposure and outcome are binary.[4] In case of a linear model and single IV, the GMM estimator is equivalent to 2SLS, additive SMM, and LIML.[14,26,52] On the other hand, with log-linear model, (i.e., MGMM),[45] it is equivalent with MSMM and provides the population CRR.[45] However, this estimator with LRM is not consistent for the causal odds ratio (COR) due to non-collapsibility of the OR.[67]

Binary outcome, continuous exposure, continuous IV

In a study of the effect of smoking (number of cigarette per day) on the risk of myocardial infarction, cigarette price can be considered as a continuous IV. Increasing the price of cigarettes has an impact on the cigarette demand per day. Hence cigarette price affects the outcome through the number of cigarettes per day.

In this case, Palmer et al.[69] suggested a two-stage IV method where the first-stage is a linear regression and the second stage-model is a logistic or log-linear model.[45] Since IV analysis with logistic regression does not provide a consistent exposure effect, in order to estimate causal risk ratio, GMM with log-linear model is preferable. It is noted that the two-stage methods incorporate with 2SLS, 2SPS, and 2SRI.

Binary outcome, binary exposure, binary instrument

In a study of influenza vaccine effectiveness on mortality risk, Yoo and Frick used a history of gout as IV.[70] The assumption underlying this study was that patients with a history of gout were more likely to visit their physician and thus have a higher likelihood of receiving the influenza vaccine, while a history of gout was assumed to be independent of other risk factors of mortality. This is an example of a study in which the outcome (mortality), the exposure (influenza vaccination) as well as the IV (history of gout) is binary.

Bivariate probit models

When the outcome of interest is binary, so-called probit models can be applied for IV analysis. In contrast to 2SLS, probit models directly model probabilities (i.e., are restricted on [0, 1]).[4,32] Bivariate probit models (BPM) can be applied in two-stages, but unlike common two-stage methods, this method is estimated via full-information maximum likelihood, which takes into account the correlation between the error terms in the two equations.[12] A more detailed model description can be found elsewhere.[4,32]

The interpretation of BPM parameters are not like those of ordinary regression model parameters (e.g., logarithm of odds ratio from a logistic model). However, by multiplying a probit coefficient by approximately 1.6 or 1.8, probit coefficients approximate the coefficients obtained through logistic regression.[4]

In case of binary outcome, linear IV methods may yield biased results and BPM may be preferable.[32,62] Furthermore, the estimates are more efficient than 2SLS, whereas 2SLS models are more robust to incorrect modelling assumptions regarding the bivariate normal distribution of the error terms.[70,71] However, when the distribution of error terms are not normal or the average probability of the outcome variable is close to one or zero, or if there is more than one endogenous exposures, the estimates from the BPM are generally not consistent for the average causal effect (ACE).[32,71]

Linear probability model

This method is a particular case of the 2SLS and provides exposure effects on the risk difference scale. When the outcome, exposure, and IV are binary, the estimator can be expressed as:

$$\hat{\beta}_{IV} = \frac{p(Y = 1 \mid Z = 1) - p(Y = 1 \mid Z = 0)}{p(X = 1 \mid Z = 1) - p(X = 1 \mid Z = 0)}$$

where $p(Y = 1|Z = 1) - p(X = 1|Z = 0)$ is the risk difference of an event between $Z = 1$ and $Z = 0$ and $p(X = 1|Z = 1) - p(X = 1|Z = 0)$ is the difference in probability of being exposed for $Z = 1$ and $Z = 0$. These probabilities can be derived from a linear model, which is then referred to as the linear probability model (LPM) and also equivalent to the RE method.[11,57,72,73]

LPM is simple to estimate and interpret as the regression coefficients based on linear regression. However, in linear IV analysis, LPM may provide ambiguous results because the common technique of linear IV is designed for a continuous response.[74] It should be noted that the LPM of binary treatment and outcome may produce predicted values outside of the 0–1 range.[30] Hence, for rare binary outcomes some predicted probabilities may become negative.[41] In addition, the probability of success increases linearly with exposure, that is, the marginal or incremental effect of exposure remains constant,[73] which is logically impossible for binary outcomes.[22]

Two-stage logistic regression

When both the outcome and exposure are binary and the interest is to use IV to estimate odds ratios (OR), two-stage logistic regression (2SLR) can be applied. 2SLR is similar to 2SLS, but instead of linear models using logistic models in both stages.[4,14] This method is fully parametric and maximum likelihood estimation (MLE) is used to estimate the parameters. If the first-stage logistic model is not correctly specified, the estimates from the second-stage can be biased.[75,76] Also, note that this method does not provide the COR.[45]

In this combination, since the IV methods with logistic regression models have many limitations,[4,45,69] LPM is widely used in epidemiological studies.[77] To estimate CRR and COR, MGMM/MSMM with log-linear model and LSMM, respectively is preferable.

Binary outcome, binary exposure, continuous instrument

In observational studies of the effects of catheterization vs. revascularization surgery on mortality risk in patients with an acute myocardial infarction, the groups of patients that reach the hospital earlier may be more likely to receive a heart catheterization. If patients that live close to a hospital are similar to those who live further away, distance to the hospital can act as an IV. In that case, IV analysis may be possible, in which the outcome (mortality) and exposure (catheterization vs. revascularization surgery) are binary, and the IV (distance to the hospital) is continuous.[78] Another example of this setting is a study of Rhythm vs. rate control treatment and all-cause mortality (yes/no) using PPP (proportion of Rhythm/ rate control treatment prescription by each physician) as a continuous IV.[77]

As mentioned in Table 2 several methods are suitable in this setting. However, assumptions of MGMM are more flexible than other methods and hence it would be a better choice.

Other outcomes

Apart from the situations discussed above, the outcome variable in epidemiologic research may also be a time-to-event or count variable. Also in case of these outcome variables, IV analysis has been applied with two-stage method. In that case, the second-stage model could be a Cox proportional hazards model.[79–81] However, Brookhart et al.[3] stated that this approach for IV analysis is not motivated by a theoretical model and, therefore, parameters that are obtained from this approach may not be causally interpretable. Examples of this approach are a study of the effect of rosiglitazone on (time to) cardiovascular hospitalization and all-cause mortality using facility-prescribing patterns as an IV,[79] and a study of the effect of adjuvant chemotherapy on (time to) breast cancer recurrence using physician preference as an IV.[80]

When dealing with count outcome variables, a two-stage IV method can be used in which the second stage model is a log-linear model.[45] However, MGMM[45] and 2SRI[17] are also applicable in this setting (Table 2).

Table 1. Overview of all reviewed IV methods (basic notions, strengths, and limitations)

IV Methods	Basic notions	Strengths	Limitations
Ratio Estimator (RE)	- the RE is appropriate when only one IV	- help to estimate the exposure effects on the outcome in easiest way and with a single binary IV and no other covariates, 2SLS = RE - for continuous outcome, the RE provides average causal effect (ACE) and for binary outcome, RE provides causal risk ratio (CRR) - for rare binary outcome, causal odds ratio (COR) is approximately equal to CRR	- RE is not consistent for the causal odds ratio - it is not suitable for multiple IVs
Two-stage Least Squares (2SLS)	- linear models assume homogeneity of variance - for multiple IVs, IV estimator is the weighted average of the ratio estimators	- natural starting point of IV analysis - the estimate asymptotically unbiased and consistent for the ACE as long as the key assumptions of IVs are satisfied - widely used for binary exposure and outcome. In that case, the exposure effect is on risk difference scale	- less efficient than traditionally adjustment methods (e.g. multivariable regression), this is also generally true for all IV methods - in small samples the bias may be substantial - may be biased results in the case of non-linear models - for multiple IVs, 2SLS estimator is biased and hence limited information of maximum likelihood method would be an alternative
Linear Probability Models (LPM)	- when the outcome, exposure, and IV are binary, the data are modelled using linear functions - exposure effect is on risk difference scale	- simple to estimate and interpret as the regression coefficients - like 2SLS, LPM estimator is also consistent for the ACE given that the assumptions are fulfilled	- sometimes predicted probabilities outside of the 0–1 range and for rare outcomes this may become negative - assumes the marginal/incremental effect of exposure remains constant which is logically impossible for binary outcome
Two-stage Predictor Substitution (2SPS)	- the rote extension to nonlinear models of the linear IV models - 2SPS is the mimic of 2SLS - non-linear least squares or any other consistent estimation technique is used to estimates the parameter	- in case of linear model, 2SPS = 2SLS and provides consistent estimator for the ACE	- in practice, 2SPS in non-linear model does not always yield consistent exposure effects on outcome

Table 1. Overview of all reviewed IV methods (basic notions, strengths, and limitations) (*Continued*)

IV Methods	Basic notions	Strengths	Limitations
Two-stage Residual Inclusion (2SRI)/Control function estimator	- include the estimated unobservable confounder (residual) from the first-stage as an additional variable along with the exposure in the second-stage model	- yields consistent estimates for linear and non-linear models and performs better than 2SPS - possible to apply in the specific case of a binary exposure with a binary or count outcome - in the case of log-linear model at stage-two, 2SRI estimator provides CRR and under the linear model, 2SRI = 2SLS = 2SPS	- bias of 2SRI estimator increases as the magnitude of confounding increases - because of non-collapsibility of the odds ratio (OR) in logistic regression model (LRM), 2SRI estimator does not provide COR
Two-stage Logistic Regression (2SLR)	- the outcome and exposure are binary and interest to estimate OR - fully parametric, maximum likelihood (ML) technique is used to estimate the parameters	- parallel to 2SLS using LRM in both stages instead of linear models	- if the first-stage logistic model is not correctly specified then second-stage parameter estimates might be biased - estimator does not provide COR
Three-stage Least Squares (3SLS)	- an extension of 2SLS but unlike the 2SLS, all coefficients are estimated simultaneously, requires three steps - in 2SLS, if the errors in the two equations are correlated, the 3SLS can be an suitable alternative	- unlike 2SLS, it uses more information of errors - provides ACE, assuming that the system of equations is correctly specified with the necessary assumptions (as in 2SLS)	- more vulnerable to a misspecification of the error terms - not widely used
Structural Mean Models (SMMs)	- SMMs use IVs via G-estimation and involves the assumption of conditional mean independence (CMI) - additive SMMs use continuous outcome and multiplicative SMMs use positive-valued outcomes - MSMM assumed log-linear model to measure the risk ratio and LSMM assumes LRM which is fitted by maximum likelihood technique and provides OR	- it relaxes several of the modelling restrictions (constant treatment effects) required by ratio estimator/two-stage methods - can be used in the case of time-dependent instruments, exposures, and confounders - additive SMMs provides average treatment effects for the treated subjects under the no effect modification by IV but consistent for the ACE under the stronger assumption of no effect modification by unmeasured confounders. Similarly, MSMM provides CRR for the treated individuals and CRR for the entire population, respectively - LSMM provides COR (treated)	- no effect modification is hard to verify in practical situations - with a binary outcome, additive SMMs and MSMM suffer from the limitations of linear and log-linear models (e.g., predicted response probabilities may outside of the interval (0, 1))

Table 1. Overview of all reviewed IV methods (basic notions, strengths, and limitations) (*Continued*)

IV Methods	Basic notions	Strengths	Limitations
Generalized Method of Moments (GMM)	- the standard IV (2SLS) estimator is a special case of a GMM estimator - a non-linear analogue of 2SLS - making assumptions about the moments of the error term - allows estimation of parameters in over-identified model (number of IV greater than number of exposure variable) - the parameters are estimated in an iterative process	- it requires specification only of certain moment conditions and consistently estimate the parameters of linear and non-linear models - non-linear GMM estimator is asymptotically more efficient than 2SLS - more robust and less sensitive to parametric conditions and works better than 2SLR when exposure and outcome are binary - in case of heteroskedasticity, this is more efficient than the linear IV estimators	- GMM estimator with logistic regression model is not consistent for the COR due to non-collapsibility of the OR
Bivariate Probit Models (BPM)	- two-stage method, but as different to 2SLS, they model probabilities directly and are restricted on [0,1] - full information ML is used to estimate the parameters and accounts for the correlation between the errors	- for binary outcome and exposure, BPM perform better than linear IV methods - the parameter estimates are asymptotically unbiased and consistent for ACE	- when the distribution of error terms are not normal or the average probability of the outcome variable is close to one or zero, the BPM estimator may not be consistent for ACE

Table 2. Decision matrix for choosing IV analytical methods for different types of outcome, exposure and IV

			IV methods¥													
Outcome	Exposure	IV	RE	2SLS	3SLS	ASMM±*	MSMM±	LSMM±	2SPS	2SRI	2-Stage Model	LGMM*	MGMM	BPM	LPM	2LRM	References of examples
Continuous	Continuous	Continuous		**X**	X				X	X		X					Hannah et al[25] Cawley J[99] Shetty et al[100]
Continuous	Continuous	Binary	**X**	X		X			X	X							Permutt and Hebel[23]
Continuous	Binary	Continuous		X					X	**X**							Cole et al[55]
Continuous	Binary	Binary	X	X		**X**			X	**X**							Ionescu-Ittu et al[41]
Binary	Continuous	Continuous#		X					X	X	X		**X**				Palmer et al[45]
Binary	Continuous	Binary	X	X			**X**		X	X	X		**X**				Ionescu-Ittu et al[77]
Binary	Binary	Continuous		X						X			**X**	X			Ionescu-Ittu et al[77] McClellan et al[78]
Binary	Binary	Binary	X	X			**X**	X		X			**X**	**X**	X	X	Brookhart et al[2,5]
Survival	Binary	Binary									X						Bosco et al[80]
Count	Binary	Binary					**X**			X	X						Mullahy J[101]

Abbreviations: RE: ratio estimator; 2SLS: two-stage least squares; 3SLS: three-stage least squares; ASMM: additive structural mean model; MSMM: multiplicative structural mean model; LSMM: logistic structural mean model; 2SPS: two-stage predictor substation; 2SRI: two-stage residual inclusion; LGMM: generalized method of moments with linear model; MGMM: multiplicative generalized method of moments; BPM: bivariate probit model; LPM: linear probability model; 2LRM: two-stage logistic regression model. X: methods possible to apply and "**X**" represents the method(s) that are best suited in respective combination. #this IV can be categorical (e.g., common homozygote, heterozygote, and rare homozygotegenotypes). ±As SMMs explicitly use counterfactuals outcomes; we propose to apply this method for binary IVs. *In case of heteroscedasticity, generalized method of moments and structural mean models are equally robust with continuous outcomes.

Standard error and characteristics of IV estimators

Consider two-stage models for IV analysis, in which the predicted value of exposure from the first-stage model is included in the second-stage model. The uncertainty around this prediction is not taken into account in the latter model, which therefore results in incorrect precision. Typically, standard errors (SEs) of the IV estimate from the second-stage model are too small.[12,32,59,60] An alternative method to estimate a correct SE is the so-called sandwich variance estimator (robust SE), which involves cross products of the predicted treatment and a dispersion factor based on the observed treatment.[29] Most statistical software packages provide this sandwich variance estimate.[21] Angrist and Krueger[21] noticed that these SEs are asymptotically valid but in practice they are only approximately valid.

An alternative way of estimating SEs is the bootstrap method.[82] Here bootstrap samples of the original data can be used to estimate the variation in the IV estimates and hence its SE.[4,6,83–85] It should be stressed that one of the weaknesses of the IV estimator is that it tends to display large SEs relative to the conventional regression estimator.[11,77] It is also noted that, the IV estimator can behave badly in finite samples and show biased results[34] and this bias is amplified when the IV is weak.[22,34]

Interpretation of estimates from IV analysis

Researchers may be interested to estimate the average treatment effects over the entire study population.[43] However, it has been argued that the basic assumptions of IV analysis are not sufficient to achieve point estimates for the causal effect of exposure on the outcome, but only estimate upper and lower bounds of this parameter.[22,86,87] To achieve a point estimate of the ACE over the entire study population, the additional strong assumption of homogeneity of exposure across levels of the IV should be satisfied.[28] Moreover, IV analysis captures the ATT under the assumption of no effect modification by IV.[28] When exposure effects are not homogeneous across IV levels, under the monotonicity assumption (i.e., the IV affects the treatment deterministically in one direction), the IV estimate quantifies the local average treatment effect (LATE),[88] which is only informative for a subset of the study population, namely those who comply with the IV.[27,43,89,90]

Assessment of IV assumptions

As noted, IV analysis must satisfy three basic assumptions and if these assumptions do not hold, results may be seriously biased.[3,11] The first assumption (i.e., the IV is related to exposure) is generally easier to check using available statistical methods than the other two assumptions. The second (IV has no direct effect on outcome) and third (IV is independent of confounders) assumptions are unverifiable or not directly testable as they involve unobservable variables.[1,9,11,18,45,70,91] Some authors proposed circumstantial evidence to support these assumptions.[2,5,92,93]

In order to check the first assumption, the F-statistic value from the first-stage linear regression model is widely used although this statistic highly affected by sample size.[70,77,84]

There is a rule of thumb that if the F-statistic value is greater than 10, the first assumption holds.[11,94,95] Other measures for the strength of the association between IV and exposure include the first-stage regression coefficient of the IV[63,96] or R^2 in linear model,[79,84,97] the odds ratio,[6,92] or pseudo-R-squared.[70] When the correlation between IV and exposure is not strong enough, IV analysis is likely to be biased (weak IV bias, which increases with the weakness of the IV). A weak IV will provide large SEs for the IV estimator.[3,11,34,62,98]

DISCUSSION

We reviewed IV methods for observational studies, highlighting their strengths and limitations for epidemiological research and developed a decision matrix (Table 2) to assist in choosing an appropriate IV method for different types of outcome, exposure and IV. These methods have some common and distinct characteristics. When the IV assumptions are violated, sample size is small and models are not correctly specified, all methods tend to perform poorly and show biased results.

The reviewed IV methods can be distinguished: moments based and semi-parametric (e.g. 2SLS, GMM, SMM) and likelihood based (e.g. BPM, 2SLR, LSMM) methods. The moment based methods or semi-parametric method are in general less efficient than likelihood based methods. However, likelihood methods are more vulnerable to incorrect modelling assumptions, in which case moment based methods are more robust. In empirical data, although several IV methods can be applicable in the same combination of IV, exposure, and outcome, considering different methods' assumptions, target parameters being estimated are different, so the interpretations of exposure effects appear different.[60] Therefore, choosing an appropriate IV method requires attention.[70]

In order to obtain ACE or LATE or ATT, along with basic assumptions, extra assumptions such as homogeneous exposure effect or monotonicity in case of heterogeneous exposure effect or no effect modification by IV, respectively should be fulfilled. These different assumptions result in the estimation of different causal effects, and hence, researchers should be aware for interpretations of the IV estimates.[22]

In RCTs, the IV of treatment assignment satisfies the assumptions by design, but in observational studies, this is not the case. In the latter situation, subject matter knowledge and theoretical motivations (why is an IV related to treatment and unrelated to patients' characteristics and outcome?) should be given especially regarding the second and third condition underlying the IV method. If the IV is weakly related to exposure and correlated with unmeasured variables, IV methods may yield biased results.[102] In addition, the main critique of any IV analysis is that the IV may affect the outcome through some pathway other than through the exposure of interest.[35] This condition cannot be verified empirically.

From a methodological perspective, the IV method is a powerful statistical tool, given that a valid IV is present and IV analysis correctly applied. In that case, it can provide a valid

estimate in the presence of measured and unmeasured confounding. However, if there is strong confounding effect, it is difficult to find an appropriate IV.[11]

A limitation of our review is that we restricted ourselves to commonly used IV methods. We did not discuss nonparametric and Bayesian IV methods. We refer to the literature for examples of the methods.[10,74,86,103–106] Because of limited space, we did not describe mathematical models with detailed derivation of IV estimators for all methods.

In conclusion, IV analysis is potentially powerful methods to control for confounding (both measured and unmeasured). Some IV methods (e.g., 2SLS, 2SRI) can be applied in many situations (different combinations of continuous and binary outcome, exposure and instrumental variables), whereas others (e.g., RE, BPM, 2SLR) can only be applied in a limited number of situations. Irrespective of the IV method that is used in a particular study, in order to provide valid interpretation of the exposure effect on the outcome, researchers should be aware of the key assumptions underlying IV methods.

REFERENCES

1. Pratt N, Roughead EE, Ryan P, Salter A. Antipsychotics and the risk of death in the elderly: An instrumental variable analysis using two preference based instruments. *Pharmacoepidemiol Drug Saf* 2010; 19(7): 699-707.
2. 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; 17(3): 268-75.
3. Brookhart MA, Rassen JA, Schneeweiss S. Instrumental variable methods in comparative safety and effectiveness research. *Pharmacoepidemiol Drug Saf* 2010; 19(6): 537-44.
4. Rassen JA, Schneeweiss S, Glynn RJ, Mittleman MA, Brookhart MA. Instrumental variable analysis for estimation of treatment effects with dichotomous outcomes. *Am J Epidemiol* 2009; 169(3): 273-84.
5. 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; 54(11): 3390-8.
6. Groenwold RHH, Hak E, Klungel OH, Hoes AW. Instrumental Variables in Influenza Vaccination Studies: Mission Impossible?! *Value Health* 2010; 13(1): 132-7.
7. Rassen JA, Brookhart MA, Glynn RJ, Mittleman MA, Schneeweiss S. Instrumental variables I: instrumental variables exploit natural variation in nonexperimental data to estimate causal relationships. *J Clin Epidemiol* 2009; 62(12): 1226-32.
8. Bennett DA. An introduction to instrumental variables-Part 2: Mendelian randomisation. *Neuroepidemiology* 2010; 35(4): 307-10.
9. Chen Y, Briesacher BA. Use of instrumental variable in prescription drug research with observational data: a systematic review. *J Clin Epidemiol* 2011; 64(6): 687-700.
10. Greenland S. An introduction to instrumental variables for epidemiologists. *Int J Epidemiol* 2000; 29(4): 722-9.
11. Martens EP, Pestman WR, De Boer A, Belitser SV, Klungel OH. Instrumental variables: Application and limitations. *Epidemiology* 2006; 17(3): 260-7.

12. Greene WH, Zhang C. Econometric analysis. 5th ed. Prentice hall New Jersey; 2003.
13. Kennedy P. A guide to econometrics. 5th ed. Cambridge, MA: MIT Press; 2003.
14. Johnston KM, Gustafson P, Levy AR, Grootendorst P. Use of instrumental variables in the analysis of generalized linear models in the presence of unmeasured confounding with applications to epidemiological research. *Stat Med* 2008; 27(9): 1539-56.
15. Foster EM. Instrumental variables for logistic regression: An illustration. *Soc Sci Res* 1997; 26(4): 487-504.
16. Terza JV, Bradford WD, Dismuke CE. The use of linear instrumental variables methods in health services research and health economics: A cautionary note. *Health Serv Res* 2008; 43(3): 1102-20.
17. Wooldridge JM. Econometric analysis of cross section and panel data. Cambridge, MA: MIT Press; 2002.
18. Didelez V, Meng S, Sheehan NA. Assumptions of IV methods for observational epidemiology. *Stat Sci* 2010; 25(1): 22-40.
19. Bennett DA. An introduction to instrumental variables analysis: Part 1. *Neuroepidemiology* 2010; 35(3): 237-40.
20. 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; 45(10): S116-22.
21. Angrist JD, Krueger AB. Instrumental variables and the search for identification: From supply and demand to natural experiments. *J Econ Perspect* 2001; 15(4): 69-85.
22. Hernán MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? *Epidemiology* 2006; 17(4): 360-72.
23. Permutt T, Hebel JR. Simultaneous-equation estimation in a clinical trial of the effect of smoking on birth weight. *Biometrics* 1989; 45(2): 619-22.
24. Wald A. The fitting of straight lines if both variables are subject to error. *The Annals of Mathematical Statistics* 1940; 11(3): 284-300.
25. Hannah S M, Cathy E E, Shengxu L, et al. Mendelian Randomisation Study of Childhood BMI and Early Menarche. *J Obes* 2011; 2011.
26. Burgess S. Statistical issues in Mendelian randomization: use of genetic instrumental variables for assessing causal associations. 2011.
27. Angrist J, Imbens G, Rubin DB. Identification of causal effects using instrumental variables *J Am Stat Assoc* 1996; 91(434): 444-72.
28. O'Malley AJ, Frank RG, Normand S-T. Estimating cost-offsets of new medications: Use of new antipsychotics and mental health costs for schizophrenia. *Stat Med* 2011; 30(16): 1971-88.
29. Wooldridge JM. Instrumental variables estimation of the average treatment effect in correlated random coefficient models. *Adv Econom* 2008; 21: 93-117.
30. Cameron AC. Microeconometrics: methods and applications. Cambridge University Press; 2005.
31. Brookhart MA, Stürmer T, Glynn RJ, Rassen J, Schneeweiss S. Confounding control in healthcare database research: Challenges and potential approaches. *Med Care* 2010; 48(6 SUPPL.): S114-20.
32. Angrist JD, Pischke JS. Mostly harmless econometrics: An empiricist's companion. Princeton Univ Pr; 2008.
33. Burgess S, Granell R, Palmer TM, Sterne JAC, Didelez V. Lack of identification in semiparametric instrumental variable models with binary outcomes. *Am J Epidemiol* 2014; 180(1): 111-9.
34. Bound J, Jaeger DA, Baker RM. Problems with instrumental variables estimation when the correlation between the instruments and the endogeneous explanatory variable is weak. *J Am Stat Assoc* 1995; 90(430): 443-50.

35. Glymour MM. Natural experiments and instrumental variable analyses in social epidemiology. In: J. Michael Oakes, Jay S. Kaufman, editor. Methods in social epidemiology. San Francisco: Jossey-Bass; 2006. p.429–460.
36. Anderson TW, Kunitomo N, Sawa T. Evaluation of the Distribution Function of the Limited Information Maximum Likelihood Estimator. *Econometrica* 1982; 50(4): pp. 1009-1027.
37. Zellner A, Theil H. Three-stage least squares: simultaneous estimation of simultaneous equations. *Econometrica* 1962; : 54-78.
38. SAS Institute Inc. SAS/ETS® 9.1. User's Guide. 2004.
39. Dowd B, Town R. Does X Really Cause Y. *Washington, DC: Academy Health, 2002*.
40. Ahn J. Beyond single equation regression analysis: Path analysis and multi-stage regression analysis. *Am J Pharm Educ* 2002; 66(1): 37-41.
41. Ionescu-Ittu R, Delaney JAC, Abrahamowicz M. Bias–variance trade-off in pharmacoepidemiological studies using physician-preference-based instrumental variables: a simulation study. *Pharmacoepidemiol Drug Saf* 2009; 18(7): 562-71.
42. Grootendorst P. A review of instrumental variables estimation in the applied health sciences. *SEDAP* 2007; Research Paper No. 215.
43. McNamee R. Intention to treat, per protocol, as treated and instrumental variable estimators given non-compliance and effect heterogeneity. *Stat Med* 2009; 28(21): 2639-52.
44. Robins JM. The analysis of randomized and non-randomized AIDS treatment trials using a new approach to causal inference in longitudinal studies. *Health service research methodology: a focus on AIDS* 1989; 113159.
45. Palmer TM, Sterne JAC, Harbord RM, et al. Instrumental variable estimation of causal risk ratios and causal odds ratios in mendelian randomization analyses. *Am J Epidemiol* 2011; 173(12): 1392-403.
46. Robins JM. Correcting for non-compliance in randomized trials using structural nested mean models. *Commun Stat* 1994; 23(8): 2379-412.
47. Maracy M, Dunn G. Estimating dose-response effects in psychological treatment trials: the role of instrumental variables. *Stat Methods Med Res* 2011; 20(3): 191-215.
48. Clarke PS, Palmer TM, Windmeijer F. Estimating structural mean models with multiple instrumental variables using the generalised method of moments. *The Centre for Market and Public Organisation* 2011.
49. Tan Z. Marginal and nested structural models using instrumental variables. *J Am Stat Assoc* 2010; 105(489): 157-69.
50. Vansteelandt S, Goetghebeur E. Causal inference with generalized structural mean models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2003; 65(4): 817-35.
51. Robins J, Rotnitzky A. Estimation of treatment effects in randomised trials with non-compliance and a dichotomous outcome using structural mean models. *Biometrika* 2004; 91(4): 763-83.
52. Clarke P, Windmeijer FAG, Centre for Market and Public Organisation (University of Bristol). Instrumental variable estimators for binary outcomes. Centre for Market and Public Organisation, University of Bristol; 2009.
53. Clarke PS, Windmeijer F. Identification of causal effects on binary outcomes using structural mean models. *Biostatistics* 2010; 11(4): 756-70.
54. Goetghebeur E, Vansteelandt S. Structural mean models for compliance analysis in randomized clinical trials and the impact of errors on measures of exposure. *Stat Methods Med Res* 2005; 14(4): 397-415.
55. Cole JA, Norman H, Weatherby LB, Walker AM. Drug copayment and adherence in chronic heart failure: Effect on cost and outcomes. *Pharmacotherapy* 2006; 26(8 I): 1157-64.

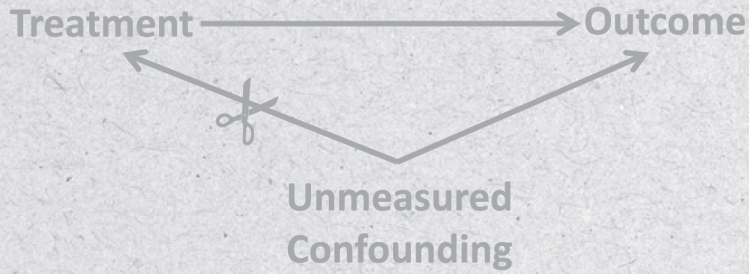
56. Lee LF. Identification and estimation in binary choice models with limited (censored) dependent variables. *Econometrica* 1979; : 977-96.
57. Galárraga O, Sosa-Rubí SG, Salinas-Rodríguez A, Sesma-Vázquez S. Health insurance for the poor: Impact on catastrophic and out-of-pocket health expenditures in Mexico. *Eur J Health Econ* 2010; 11(5): 437-47.
58. Terza JV, Basu A, Rathouz PJ. Two-stage residual inclusion estimation: Addressing endogeneity in health econometric modeling. *J Health Econ* 2008; 27(3): 531-43.
59. Cai B, Small DS, Have TRT. Two-stage instrumental variable methods for estimating the causal odds ratio: Analysis of bias. *Stat Med* 2011; 30(15): 1809-24.
60. Palmer TM, Lawlor DA, Harbord RM, et al. Using multiple genetic variants as instrumental variables for modifiable risk factors. *Stat Methods Med Res* 2011; 21(3): 223-42. doi:10.1177/0962280210394459.
61. Hausman JA. Specification tests in econometrics. *Econometrica* 1978; : 1251-71.
62. Suh HS. The effect of using fibrates in conjunction with statins for the management of dyslipidemia in persons with type II diabetes mellitus.[dissertation]. UNIVERSITY OF SOUTHERN CALIFORNIA; 2009.
63. Stuart BC, Doshi JA, Terza JV. Assessing the impact of drug use on hospital costs. *Health Serv Res* 2009; 44(1): 128-44.
64. Newey WK. Efficient estimation of limited dependent variable models with endogenous explanatory variables. *J Econ* 1987; 36(3): 231-50.
65. Hansen LP. Large sample properties of generalized method of moments estimators. *Econometrica* 1982; : 1029-54.
66. Baum CF, Schaffer ME, Stillman S. Instrumental variables and GMM: Estimation and testing. *Stata J* 2003; 3(1): 1-31.
67. Clarke P, Windmeijer F. Instrumental Variable Estimators for Binary Outcomes. 2010; Working Paper 10/239,Centre for Market and Public Organisation, Univ. Bristol.
68. Lee L-. GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. *J Econ* 2007; 137(2): 489-514.
69. Palmer TM, Thompson JR, Tobin MD, Sheehan NA, Burton PR. Adjusting for bias and unmeasured confounding in Mendelian randomization studies with binary responses. *Int J Epidemiol* 2008; 37(5): 1161-8.
70. Yoo BK, Frick KD. The instrumental variable method to study self-selection mechanism: A case of influenza vaccination. *Value Health* 2006; 9(2): 114-22.
71. Bhattacharya J, Goldman D, McCaffrey D. Estimating probit models with self-selected treatments. *Stat Med* 2006; 25(3): 389-413.
72. Angrist JD. Estimation of limited dependent variable models with dummy endogenous regressors: Simple strategies for empirical practice. *J Bus Econ Stat* 2001; 19(1): 2-16.
73. Carrasco R. Binary choice with binary endogenous regressors in panel data: Estimating the effect of fertility on female labor participation. *J Bus Econ Stat* 2001; 19(4): 385-94.
74. Baiocchi M, Rosenbaum P, Small D, Lorch S. Near/Far Matching-A Nonparametric Instrumental Variables Technique for Binary Outcomes. 2009.
75. Angrist JD. Estimation of limited dependent variable models with dummy endogenous regressors. *J Bus Econ Stat* 2001; 19(1): 2-28.
76. Henneman TA, Van Der Laan MJ, Hubbard AE. Estimating causal parameters in marginal structural models with unmeasured confounders using instrumental variables. *(U.C. Berkeley Division of Biostatistics Working Paper Series, paper 104). Berkeley, CA: The Berkeley Electronic Press* 2002.

77. Ionescu-Ittu R, Abrahamowicz M, Pilote L. Treatment effect estimates varied depending on the definition of the provider prescribing preference-based instrumental variables. *J Clin Epidemiol* 2012; 65(2): 155-62.
78. McClellan M, McNeil BJ, Newhouse JP. Does More Intensive Treatment of Acute Myocardial Infarction in the Elderly Reduce Mortality? *JAMA* 1994; 272(11): 859-66.
79. Ramirez SPB, Albert JM, Blayney MJ, et al. Rosiglitazone is associated with mortality in chronic hemodialysis patients. *Journal of the American Society of Nephrology* 2009; 20(5): 1094-101.
80. Bosco JLF, Silliman RA, Thwin SS, et al. A most stubborn bias: no adjustment method fully resolves confounding by indication in observational studies. *J Clin Epidemiol* 2010; 63(1): 64-74.
81. Schmoor C, Caputo A, Schumacher M. Evidence from nonrandomized studies: A case study on the estimation of causal effects. *Am J Epidemiol* 2008; 167(9): 1120-9.
82. Efron B, Tibshirani R. An introduction to the bootstrap. Chapman & Hall/CRC; 1993.
83. Cain LE, Cole SR, Greenland S, Brown TT, Chmiel JS, Kingsley L, Detels R. Effect of highly active antiretroviral therapy on incident AIDS uising calendar period as an instrumental variable. *Am J Epidemiol* 2009; 169(9): 1124-32.
84. 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; 19(4): 1064-70.
85. Finke R, Theil H. Bootstrapping for standard errors of instrumental variable estimates. *Econ Lett* 1984; 14(4): 297-301.
86. Balke A, Pearl J. Bounds on treatment effects from studies with imperfect compliance. *J Am Stat Assoc* 1997; 92(439): 1171-6.
87. Robins JM, Greenland S. Identification of causal effects using instrumental variables: comment. *Journal of the American Statistical Association* 1996; 91(434): 456-8.
88. Imbens GW, Angrist JD. Identification and estimation of local average treatment effects. *Econometrica* 1994; : 467-75.
89. Brooks JM, Fang G. Interpreting treatment-effect estimates with heterogeneity and choice: simulation model results. *Clin Ther* 2009; 31(4): 902-19. doi:10.1016/j.clinthera.2009.04.007.
90. Fang G, Brooks JM, Chrischilles EA. Apples and oranges? Interpretations of risk adjustment and instrumental variable estimates of intended treatment effects using observational data. *Am J Epidemiol* 2011; 175(1): 60-5. doi:10.1093/aje/kwr283.
91. Didelez V, Sheehan N. Mendelian randomization as an instrumental variable approach to causal inference. *Stat Methods Med Res* 2007; 16(4): 309-30.
92. 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. *Can Med Assoc J* 2007; 176(5): 627.
93. Brookhart MA, Schneeweiss S. Preference-based instrumental variable methods for the estimation of treatment effects: Assessing validity and interpreting results. *Int J Biostat* 2007; 3(1): 14.
94. Stock JH, Wright JH, Yogo M. A survey of weak instruments and weak identification in generalized method of moments. *J Bus Econ Stat* 2002; 20(4): 518-29.
95. Sheehan NA, Meng S, Didelez V. Mendelian randomisation: a tool for assessing causality in observational epidemiology. *Methods Mol Biol* 2011; 713: 153-66.
96. 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.

97. Rassen JA, Brookhart MA, Glynn RJ, Mittleman MA, Schneeweiss S. Instrumental variables II: instrumental variable application-in 25 variations, the physician prescribing preference generally was strong and reduced covariate imbalance. *J Clin Epidemiol* 2009; 62(12): 1233-41.
98. Staiger D, Stock JH. Instrumental variables regression with weak instruments. *Econometrica* 1997; 65(3): 557-86.
99. Cawley J. The impact of obesity on wages. *J Hum Resour* 2004; 39(2): 451-74.
100. Shetty KD, Vogt WB, Bhattacharya J. Hormone replacement therapy and cardiovascular health in the United States. *Med Care* 2009; 47(5): 600.
101. Mullahy J. Instrumental-variable estimation of count data models: Applications to models of cigarette smoking behavior. *Rev Econ Stat* 1997; 79(4): 586-93.
102. Hahn J, Hausman J. A new specification test for the validity of instrumental variables. *MIT Department of Economics Working Paper No.99-11* 2000.
103. Burgess S, Thompson SG, Andrews G, et al. Bayesian methods for meta-analysis of causal relationships estimated using genetic instrumental variables. *Stat Med* 2010; 29(12): 1298-311.
104. McKeigue PM, Campbell H, Wild S, et al. Bayesian methods for instrumental variable analysis with genetic instruments ('Mendelian randomization'): example with urate transporter SLC2A9 as an instrumental variable for effect of urate levels on metabolic syndrome. *Int J Epidemiol* 2010; 39(3): 907-18.
105. Gao C, Lahiri K. A comparison of some recent Bayesian and classical procedures for simultaneous equation models with weak instruments. *Department of Economics, SUNY Albany* 1999.
106. Kleibergen F, Zivot E. Bayesian and classical approaches to instrumental variable regression. *J Econ* 2003; 114(1): 29-72.

Unmeasured confounding IV PERR SCCS Observational studies
PROTECT Causal Inference Statistical methods পরিসংখ্যানগত পদ্ধতি
Statistische methoden IV PERR SCCS Pseudo Randomization
Pharmacoepidemiology IV Analyses PERR SCCS
General Practice Databases Residual Confounding
Electronic Healthcare Databases Epidemiology
Time-fixed Confounding Time-fixed Exposure
Time-varying Exposure
Time-varying Confounding
Non-randomized Studies
IV PERR SCCS
PROTECT 2SLS
IV IV

CHAPTER III
SIMULATION STUDIES OF INSTRUMENTAL VARIABLE ANALYSIS



CHAPTER 3.1

Performance of Instrumental Variable Methods in Cohort and Nested Case-Control Studies: A Simulation Study

Md Jamal Uddin, Rolf HH Groenwold, Anthonius de Boer, Svetlana V Belitser, Kit CB Roes, Arno W Hoes, Olaf H Klungel

Pharmacoepidemiology and Drug Safety 2014; 23:165–177

ABSTRACT

BACKGROUND: Instrumental variable (IV) analysis is becoming increasingly popular to adjust for confounding in observational pharmacoepidemiological research. One of the prerequisites of an IV is that it is strongly associated with exposure; if it is weakly associated with exposure, IV estimates are reported to be biased. We aimed to assess the performance of IV estimates in various pharmacoepidemiological settings.

METHODS: Data were simulated for continuous/binary exposure, outcome, and IV in cohort and nested case-control (NCC) designs with different incidences of the outcome. Pearson's correlation (PC), point bi-serial correlation, odds ratio (OR), and F-statistic were used to assess the IV-exposure association. Two-stage analysis was performed to estimate the exposure effect.

RESULTS: For all types of IV and exposure in the cohort and NCC designs, IV estimates were extremely unstable and biased when the IV was very weakly associated with exposure (e.g. PC < 0.15 for continuous or OR < 2.0 for binary IV and exposure; although specific cut-off values depend on simulation settings). For stronger IVs, estimates were unbiased and become less variable compared to weaker IVs in the case of continuous and binary (risk difference scale) outcomes. For a similar IV-exposure association (e.g. OR = 1.4 and 5% incidence of the outcome), the variability of the estimates was more pronounced in the NCC (SD = 2.37, case: control = 1:5) compared to the cohort design (SD = 1.14). The variability was even more pronounced for rare (≤ 1%) outcomes. However, IV estimates from the NCC design became less variable with an increasing number of controls per case. Moreover, estimates were biased when the IV was related to confounders even with strong IVs.

CONCLUSIONS: IV analysis performs poorly when the IV-exposure association is extremely weak, especially in the NCC design. IV estimates in the NCC design become less variable when the number of control increases. As NCC does not use the entire cohort, in order to achieve stable estimates, this design requires a stronger IV-exposure association than the cohort design.

INTRODUCTION

Instrumental variable (IV) analysis is becoming increasingly popular to adjust for confounding in observational pharmacoepidemiological research.[1–10] An IV is a variable that is associated with the exposure under study, and only related to the outcome through exposure. Hence, an IV should neither directly nor indirectly (e.g. through confounders) be associated with the outcome. In that case, IV analysis controls for measured and unmeasured confounding and provides consistent and asymptotically unbiased estimates of the exposure on the outcome.[5,11–14]

However, when the association between an IV and exposure is weak, the IV itself is called a weak instrument or weak IV.[15] In that situation, the statistical models of IV analysis are "weakly identified",[16] and the exposure effects obtained by IV analysis are often inconsistent and biased with wide confidence intervals.[5,11,17–21]

Over the last two decades there have been a number of studies[5,8–11,14,16,18,22–28] that have addressed the impact of weak IV(s). These articles focused, for example, on inconsistent and biased estimates, small sample bias, the impact of weak IVs on confidence intervals in the case of continuous or binary outcomes, and all focused on a cohort design only. These studies did not cover the entire range of common (pharmacoepidemiological settings, e.g. cohort as well as (nested) case-control (NCC) designs, different types of exposure, IV, and outcome (i.e., both continuous and binary), and different incidences of the outcome (especially rare outcomes), and slight deviations of the IV assumptions due to imbalances of confounder distributions between IV levels. The aim of the present study was to assess the performance of IV estimates in different realistic pharmacoepidemiological settings using simulated data.

METHODS

We used simulated data to assess the performance of IV estimates in a cohort and a nested case-control design. Different combinations of continuous or binary IV, exposure, and outcome were examined in both designs (obviously in the NCC design only a binary outcome). The following notations were used: Y denotes the outcome, X denotes the exposure/treatment, Z denotes the IV, and C denotes a set of confounding variables (some possibly unmeasured). For simplicity, we assumed time-independent exposure, IV, and confounders. The simulation settings are presented in Table 1. We used statistical software R (Windows, version 2.15.1) to simulate and analyse our data.[29]

Simulation of the data

Step 1: Basic setup

We studied several sample sizes in both designs. For the cohort data, sample sizes were 1000, 5000, and 10,000. For the NCC data, samples were drawn from large cohorts with sizes of 10,000 and 50,000. The incidences of the outcome we studied were 1%, 5%, 10%, and 25% for the cohort and 1% and 5% for the cohorts from which the cases and controls in the NCC design were extracted. All scenarios were simulated 10,000 times.

We imposed the following restrictions to the simulated data in order to meet the assumptions underlying IV analysis: the IV was independent of confounders; the IV was independent of the outcome given exposure and confounders; and in order to identify a point estimate of exposure on the outcome (such as the average causal effect (ACE)), the effect of exposure on the outcome was the same for all subjects (homogeneous exposure effect).

Step 2: Generation of IV and confounding variables

We assumed that the IV and confounding factor followed a multivariate normal distribution. For simplicity, we simulated a single confounding variable. For a valid IV, the correlation between IV and confounder was zero (*PC* = 0). We checked this in our simulated data by assessing the empirical association between these variables and observed that the simulated IV was indeed uncorrelated with the confounder. In a separate simulation, we also assessed the impact of violation of this assumption, by imposing a correlation between the IV and the confounding factor (*PC* = 0.10 and 0.40). In the case of a binary IV, the continuous IV was dichotomized to create binary variables.[30] A cut-off value was used for dichotomization that resulted in a prevalence of the binary IV of 40%.

It is noted that in pharmacoepidemiological studies the use of IVs is still rare but it has increased over the last years.[31] One of the most often used IVs in pharmacoepidemiology is physician preference.[32] This IV can be operationalized in different ways, e.g., by using the previous prescription choice (for e.g., drug A or drug B) made by a physician, or by using the proportion of prescriptions (e.g., of drug A vs. B) by a certain physician. This proportion is a continuous variable with values in the range (0, 1), which could follow approximately a truncated normal distribution. We therefore assumed that our simulated binary IV can be comparable with physician prescribing preferences based on the previous prescription, whereas the continuous IV can be comparable with the proportion of a prescription for a certain drug.

Although in the pharmacoepidemiological studies binary IVs are mostly used, one can imagine continuous IVs in pharmacoepidemiological studies. For example, in a study of COX-2 inhibitor use, physician preference (i.e., the historical proportion of a physician's NSAID prescriptions that were for a COX-2 inhibitor) was used as an IV.[1] Another example is the amount of drug co-payment by patients, which served as a (continuous) IV in a study of the effects of adherence to beta-blocker therapy on blood pressure).[33] These IVs approximately follow a normal distribution and we therefore assumed that our continuous IV is a realistic IV in the pharmacoepidemiological settings.

Step 3: Generation of exposure variables

The continuous and binary exposures were generated based on a linear model (see below, equation 1) and logistic model (equation 2), respectively.

$$\mathrm{X} = \beta_0 + \beta_z \mathrm{Z} + \beta_c \mathrm{C} + \varepsilon; \quad [1]$$

$$logit[Prob(X=1 \mid Z,C)] = \beta_0 + \beta_z \mathrm{Z} + \beta_c \mathrm{C} \text{ and } X \sim Bernoulli\ (p); \quad [2]$$

where Z indicates the IV generated in *step 2*, the variable *C* denotes the confounding factor, β_0, β_z, and β_c denote the intercept, IV, and confounder effects on the exposure, respectively. In equation 1, ε is the error term for the exposure, which follows a standard normal

distribution (mean zero, variance 1). *p* is the probability of exposure, from the logistic model in equation 2.

Table 1. Overview of simulation settings

	Scenarios
Exposure (*X*)	Continuous: X ~ $N(0,\sigma^2)$ Binary: X ~ *Bernoulli (p)*; Prevalence (*Px*) = 0.50
Outcome (*Y*)	Continuous: Y ~ $N(0,\sigma^2)$ Binary: Y ~ *Bernoulli (p)* Incidence of Y in the cohort = 1%, 5%, 10%, & 25% Incidence of Y in the NCC = 1% & 5%
Instrumental variable (*Z*)	Continuous: Z ~ *N*(0,1) Binary: Z ~ *Bernoulli (p)*; Prevalence (P_z) = 0.40
Confounding factors	Continuous: C ~ *N*(0,1) Confounding effects: β_c = 0.50 to 2.0 Confounding effects (RD model): β_c = 0.005
Correlation between exposure and IV	PC: 0.01 to 0.60 (both X and Z are continuous) PBC: 0.01 to 0.60 (X continuous and Z binary or vice versa) OR: 1.0 to 6.0 (both X and Z are binary)
Correlation between IV and confounder	PC = 0.00 (i.e. IV is independent of confounders, valid IV) PC = 0.10 and 0.40 (i.e. IV is not independent of confounders, invalid IV)
Combination of exposure and IV either outcome is continuous or binary	continuous exposure and continuous IV continuous exposure and binary IV binary exposure and continuous IV binary exposure and binary IV
Sample size (*n*)	Cohort: 1000, 5000, and 10000 NCC: cohort sizes are 10000 and 50000
Number of simulations (*n.sim*)	10000
Case: Control in the NCC design	1:1, 1:5, and 1:10
True exposure effect	Continuous outcome: $\beta_x = 1$ Binary outcome (OR scale): $\beta_x = ln(2)$ and $\beta_x = ln(1)$, i.e. OR = 2 and OR = 1, respectively Binary outcome (RD scale): β_x = 0.005 Intercept: β_{0x} = 0.10 to 1.0 (continuous exposure) β_{0x} = -1.0 to 1.0 (binary exposure)
Intercepts of the outcome models	Continuous outcome: β_{0Y} = -1 Binary outcome (OR scale): β_{0Y} = -1 to -6 Binary outcome (RD scale): β_{0Y} = 0.0998
Nominal coverage probability	0.95

PC: Pearson's correlation coefficient; PBC: point bi-serial correlation; OR: odds ratio; NCC: nested case-control; IV: instrumental variable; RD: risk difference

In order to assess the impact of the strength of the association between IV and exposure, the value of β_z was varied over a range of values. In addition, to assess the impact of various confounding effects, different values of β_c were considered (details in Table 1).

Step 4: Generation of outcome variables

The continuous and binary outcomes (estimates in odds ratio (OR) and risk difference (RD) scales) were generated by using the following models where equation (3) is a linear model (continuous outcome), equation (4) is a logistic model (OR scale), and equation (5) is a linear RD model (RD scale).

$$\mathrm{Y} = \beta_0 + \beta_x \mathrm{X} + \beta_c \mathrm{C} + \varepsilon; \quad [3]$$

$$logit\left[Prob(Y=1 \mid X,C)\right] = \beta_0 + \beta_x \mathrm{X} + \beta_c \mathrm{C} \text{ and } Y \sim Bernoulli\ (p); \quad [4]$$

$$Prob(Y=1 \mid X,C) = \beta_0 + \beta_x \mathrm{X} + \beta_c \mathrm{C} \text{ and } Y \sim Bernoulli\ (p) \quad [5]$$

where X indicates the exposure variable generated in *step 3*, the variable *C* denotes the confounding factor generated in step 2, β_0 and β_x denote the intercept and true exposure effect on the outcome, respectively, β_c denotes the effect of the confounder on the outcome. In the linear model (3), ε is the error term for the outcome, which follows a normal distribution with mean zero and variance 1. *p* is the probability of the outcome, based on the models (4) as well as (5). Since bias of the IV estimate is invariant to the value of the parameter β_x, the exposure effects were $\beta_x = 1$ for the continuous outcome, $\beta_x = \log(2)$ and $\beta_x = 0.005$ for the binary outcome on the OR and RD scale, respectively.

Step 5: Study designs

IV analysis in a cohort design is relatively well known in pharmacoepidemiological studies. The case-control design is popular in pharmacoepidemiological studies, which is a more efficient design than a cohort design, particularly in situations where outcomes are rare or information on key variables is hard or expensive to obtain (e.g. study on gene expression data). Therefore, we also evaluated performance of IV estimates in a nested case-control (NCC) design.

A NCC study is a case-control study that is nested within a larger cohort of known size; hence, the sampling fraction of cases and controls is known. In this setting, validity of the IV analysis may be equivalent to that of IV analysis in a cohort study, given that the IV is valid and strongly associated with the exposure and the control subjects are sampled appropriately.[34]

We simulated data for the NCC design in the following way. Firstly, a large cohort with 1% and 5% incidence of the outcome was generated. Secondly, all cases were selected from the cohort data and control subjects were randomly selected from that cohort (a cross-

sectional approach, which is valid given the low cumulative incidence of the outcome over the study period). Although matching in NCC studies is typically done in order to increase efficiency of control for confounding, we did not consider matching, because we aimed to control for confounding by means of IV analysis. We considered a ratio of case to control of 1:1, 1:5, and 1:10. This means that, for example, in case of a cohort of 10,000 subjects in which the case-control study is nested, and an incidence of the outcome of 1% with 10 controls per each case, the effective sample size for the IV analysis based on NCC data is 1100 subjects.

Analysis of simulated data

Cohort data: We analysed the data using a two-stage IV method. In all settings, the first-stage model was a linear regression model, irrespective of whether the exposure was continuous or binary.[19,35] In this model, the exposure was the dependent and the IV was the independent variable.

The second-stage model was a linear regression model in the case of a continuous outcome (equation 6, below) as well as in case of a binary outcome (IV estimates on RD scale, equation 7), and a logistic regression model[36] (LRM) in the case of a binary outcome (IV estimates on OR scale, equation 8). In the second-stage model, the dependent variable was the outcome and the independent variable was the predicted value of the exposure (obtained from the first-stage model) rather than the actual exposure. It should be noted that in all settings of IV analysis, the confounder "C" was considered as unmeasured, so that in all analyses the variable "C" was omitted from the IV models.

The statistical models for the second-stage are given in equations (6), (7), and (8).

Second-stage (continuous outcome): $Y_i = \beta_0 + \beta_{IV}\hat{X}_i + \varepsilon_i$; for $i = 1,2,.....n$ [6]

Second-stage (binary outcome, RD): $\text{Prob}(Y = 1) = \beta_0 + \beta_{IV}\hat{X}_i$; for $i = 1,2,.....n$ [7]

Second-stage (binary outcome, OR): $logit(p_i) = \beta_0 + \beta_{IV}\hat{X}_i$; for $i = 1,2,.....n$ [8]

where $\hat{X}_i$ denotes the predicted value of the exposure, for binary exposure (X), the predicted value of X represents the probability of X = 1 (conditional on IV) estimated from the first-stage IV model, Y denotes the continuous outcome, and p_i is the probability of having the binary outcome. ε_i follows a normal distribution with mean zero and constant variance σ^2. The regression coefficient (β_{IV}) estimated from the second-stage model denotes the IV estimator. For binary outcome, IV estimates were on the RD scale and OR scale (equations 7 and 8), respectively. As a comparison, we also estimated exposure effects using conventional models (regression of the outcome on exposure), in which the confounder was considered unmeasured.

The IV estimators, $\hat{\beta}_{IV}$, from the equation (6 and 7) provide estimates of the ACE of the exposure for all subjects of the study population given that the assumptions of IV were fulfilled.[12] However, $\hat{\beta}_{IV}$ from the equation (8) does not generally provide a consistent estimate of the causal odds ratio (COR).[36] In order to achieve consistent estimates with LRM, we also simulated data under the "null" hypothesis,[12] i.e. the exposure effect was set to zero (OR=1).

NCC data: Data were analysed in a similar way in the NCC design with the first-stage model weighted by the inverse of the sampling fraction of cases (i.e., 1/1 = 1) and controls (i.e., 1/ (no. controls / [cohort size – no. cases]) = [cohort size – no. cases] / number of controls). The second-stage model was the (unweighted) LRM given in equation (8).

Strength of the IV

Different measures were used to assess the strength of the IV-exposure association. In order to assess the association between a continuous IV and a continuous exposure, the Pearson's correlation (PC) was used. When a binary IV and continuous exposure (or vice versa) were present, the point bi-serial correlation (PBC)[37,38] was used, and in case of a binary exposure and binary IV the odds ratio (OR) was used. Additionally, the strength of the IV was also verified by the partial F-statistic value of the first-stage regression model although this statistic is highly affected by the sample size. Throughout the article we refer to this as the F-statistic.

Bias, standard error, root mean square error, and coverage probability

Each scenario was simulated 10,000 times. Bias of IV estimates was defined as the difference between the mean of IV estimates based on 10,000 simulation runs, and the true exposure effect. Two types of confidence intervals (CI) were estimated and reported: 1.) to identify the precision of estimating the bias of the IV estimates; and 2.) to assess the variability between estimates from 10,000 simulation runs (i.e., the variation between different studies). The first CIs were estimated using the standard errors of the mean of the estimates (i.e., standard deviation of the IV estimates divided by square root of the number of simulations) and the second CIs were estimated by 2.5 and 97.5 percentiles of the 10,000 estimates.

In case of a binary outcome, we also estimated bias with no treatment effect (OR = 1), to achieve consistent estimates and evaluate the impact of non-collapsibility[39] of the ORs in the LRM (second-stage model for binary outcomes, equations 8). Since the second-stage model (equation 8) is not conditional on the confounder, it does not estimate a conditional exposure effect. As a comparison for the IV estimates we also estimated marginal exposure effects based on marginal structural models (MSMs) in the cohort design.[40] For these MSMs, the confounding factor (C) was considered a measured variable. As mentioned above, the data generation process was under a homogeneous treatment effect, which allows for a comparison between IV estimates and MSMs estimates.[41] In the MSMs, inverse probability of treatment weighting (IPTW) was applied to estimate the marginal treatment effect, including the measured confounder in the treatment model. For a binary exposure, the

IPTW was estimated by logistic regression and stabilized. For a continuous exposure, the IPTW was estimated by density functions of a linear model.[40] We assessed the sensitivity of the estimates to weight truncation at 0.5 and 99.5 percentiles, at 1.25 and 98.75 percentiles, and at 2.5 and 97.5 percentiles.[42] To evaluate the performance of IV estimates in the context of bias, accuracy, and coverage for different scenarios, root mean square error (RMSE) and coverage probability were estimated for settings with a continuous outcome.[30] The 95% coverage probability was estimated based on CIs that were estimated by 2.5 and 97.5 percentiles of the estimates from 1,000 bootstrap samples in each simulation run.

RESULTS

In our simulations, results from conventional analyses were biased due to unmeasured confounding. For example, the association between a continuous exposure and a continuous outcome was estimated to be 1.50 to 1.24 (for different settings) using conventional linear regression model instead of a true exposure effect of 1 (Table 2). Likewise, the association between a binary exposure and a binary outcome was estimated to be OR 4.01 to OR 3.72 (for different settings) using conventional LRM instead of a true exposure effect of OR = 2.0 (Table A1, Appendix).

Figure 1 shows that in a cohort design with a continuous outcome, the IV estimates were highly unstable and also be biased if the association between IV and exposure was extremely weak. For example, in the case of both a continuous IV and exposure or both a binary IV and exposure and a cohort size of 10,000, the IV estimates were biased if the correlation between IV and exposure was smaller than 0.15 or the odds ratio was smaller than 2.0, respectively. The specific cut-off values differed between simulation settings. Similar patterns were observed for other types (binary and continuous) of IV and exposure. Although the bias was within 5% of the true exposure effect of 1, this magnitude depends on simulation settings.

Furthermore, the variation in IV estimates between different simulations increased with weaker associations between IV and exposure (details in appendix, Figure A1). This is also reflected by the large RMSE in case of a weak IV-exposure association (Table 2). Both bias and variability of IV estimates increases with decreasing strength of the IV-exposure association (Table 2). In case of an IV that is not weakly associated with exposure (and hence no bias), the coverage rates of IV analysis were indeed close to the nominal level. However, for extremely weak IV-exposure associations, the coverage rate exceeded the nominal level, due to the increased standard errors for weak IV. Moreover, in this case, the mean width of the 95% CI's (calculated across 10,000 simulated samples) was very large (Table 2).

Table 2. Estimates, RMSE and coverage probability of simulation of a continuous outcome in the cohort design

Effect of IV on exposure (β_z)	IV continuous X continuous Y continuous							IV binary X continuous Y continuous						
	IV Method ¥					Conventional Method**		IV Method					Conventional Method	
	PC	Estimate	RMSE	Coverage* Probability	Width CI[a]	Estimate	RMSE	PBC	Estimate	RMSE	Coverage Probability	Width CI	Estimate	RMSE
0.01	0.01	2.201	179.136	0.989	20.270	1.500	0.500	0.00	1.317	65.772	0.992	21.793	1.500	0.500
0.05	0.04	0.930	2.167	0.962	2.686	1.499	0.500	0.02	1.018	12.234	0.980	12.966	1.500	0.500
0.10	0.07	0.991	0.148	0.947	0.608	1.498	0.498	0.04	0.937	0.824	0.962	2.785	1.499	0.500
0.15	0.11	0.996	0.096	0.947	0.383	1.494	0.495	0.05	0.979	0.213	0.950	0.946	1.499	0.499
0.60	0.39	1.000	0.024	0.947	0.092	1.424	0.424	0.20	0.999	0.048	0.949	0.189	1.479	0.479
1.50	0.73	1.000	0.009	0.947	0.037	1.235	0.235	0.46	1.000	0.019	0.949	0.075	1.394	0.394
	IV continuous X binary Y continuous							IV binary X binary Y continuous						
	IV Method					Conventional Method		IV Method					Conventional Method	
	PBC	Estimate	RMSE	Coverage Probability	Width CI	Estimate	RMSE	OR	Estimate	RMSE	Coverage Probability	Width CI	Estimate	RMSE
0.01	0.00	2.476	160.125	0.997	66.228	1.826	0.827	1.00	12.946	1505.620	0.996	68.099	1.826	0.827
0.05	0.02	1.144	33.540	0.978	27.578	1.826	0.827	1.04	2.110	162.576	0.992	54.763	1.826	0.827
0.10	0.04	0.944	0.836	0.952	4.266	1.825	0.826	1.09	0.753	21.343	0.978	28.370	1.826	0.826
0.15	0.05	0.976	0.481	0.949	2.016	1.824	0.824	1.13	0.675	16.677	0.963	10.728	1.826	0.826
0.60	0.20	0.998	0.120	0.949	0.471	1.785	0.786	1.64	0.994	0.238	0.949	0.937	1.816	0.816
1.50	0.46	1.000	0.059	0.949	0.229	1.643	0.644	3.48	0.999	0.096	0.949	0.376	1.762	0.763

IV: instrumental variable; X: exposure; Y: outcome; PC: Pearson's correlation; PBC: point bi-serial correlation; OR: odds ratio; RMSE: root mean square error, sample size: n = 10000 and number of simulations = 10000; ¥IV estimate from two-stage linear IV models and the IV is independent of the confounders (valid IV); *Nominal coverage probability is 0.95 and the coverage probabilities for conventional model were almost close to zero;**Conventional method: linear regression model; [a]Width of CI: Mean width of the 95% CI's (calculated across 10,000 simulated samples).

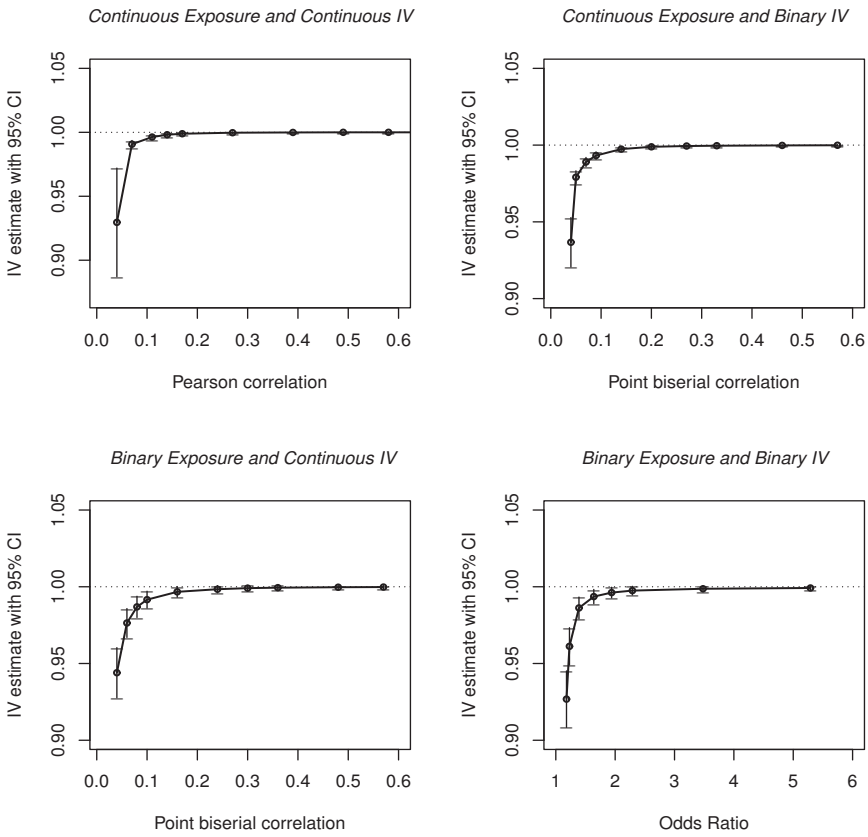


Figure 1. Mean of IV estimates from simulations of a cohort design with a continuous outcome

X-axis represents the association between exposure and IV. The horizontal straight line (dotted line) represents the reference line (true exposure effect = 1). Vertical bars indicate 95% confidence intervals around the mean of the estimates. Results are based on simulations of a cohort with sample size 10,000, and each scenario was simulated 10,000 times.

In case of a binary outcome that was simulated based on a logistic model in the cohort design, a similar pattern was observed as for the situation with a continuous outcome: a weak IV-exposure association resulted in biased IV estimates (Figure 2). In addition, estimates were also systematically biased for strong IV-exposure associations in case of a LRM for the second-stage of IV analysis (Figure 2). This is partly due to non-collapsibility of the odds ratio, but it can also be a result of model misspecification. Moreover, in case of a non-zero exposure effect comparing the IV estimates with estimates from MSMs, the IV estimates for binary exposure were closer to the MSM estimates than to the conditional effects from the data generating model (Table A1, Appendix). For continuous exposure, we examined the MSM estimates with and without truncation of weights. In the first case, the estimates were stable, whereas in the latter, the MSM estimates were highly unstable due to extreme weights and far from the IV estimates

(Table A1, Appendix). Moreover, we assessed the sensitivity of exposure effect estimates to weight truncation and we observed that the estimates were more close to IV estimates when 1% of the extreme weights were truncated (truncation at 0.5 and 99.5 percentiles) than 2.5% or 5% (truncation at 1.25 and 98.75 percentiles and at 2.5 and 97.5 percentiles, respectively). In all cases, the RMSE of MSM estimates was lower than the RMSE of IV estimates.

There was no evidence of bias for the IV estimates on RD scale, as the 95% CI's for the mean estimate always included the true exposure effect (Figure 3). The variability of the estimates in the case of extremely weak IVs was the same as described before for other settings (data not shown).

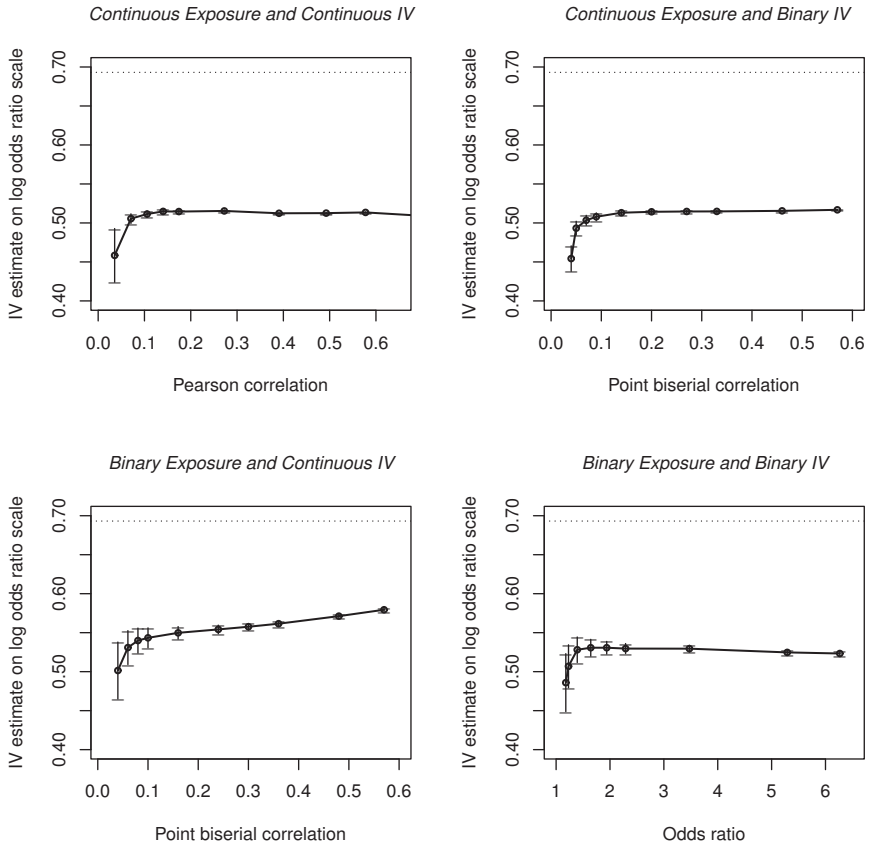


Figure 2. Mean of IV estimates from simulations of a cohort design with a binary outcome, using a logistic model in the second-stage of IV analysis

X-axis represents the association between exposure and IV. The horizontal straight line (dotted line) represents the reference line (true exposure effect = log(2)). Vertical bars indicate 95% confidence intervals around the mean of the estimates. Results are based on simulations of a cohort with sample size 10,000, incidence of the outcome 10%, and each scenario was simulated 10,000 times.

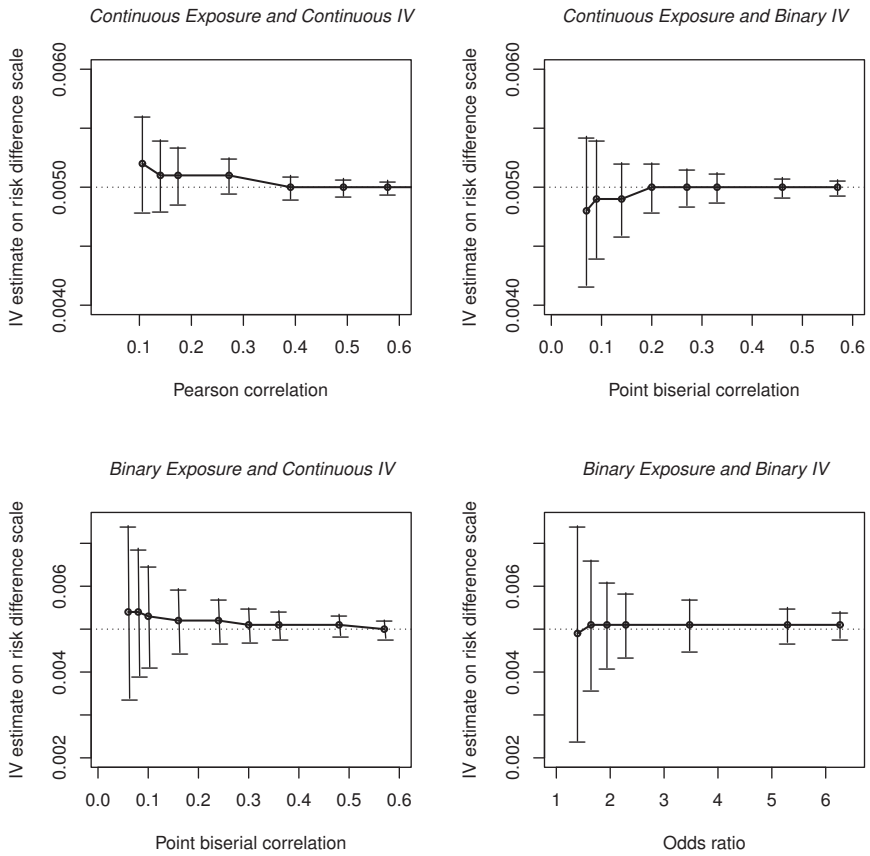


Figure 3. Mean of IV estimates from simulations of a cohort design with a binary outcome, using a linear model in the second stage of IV analysis

X-axis represents the association between exposure and IV. The horizontal straight line (dotted line) represents the reference line (true exposure effect on RD scale = 0.005). Vertical bars indicate 95% confidence intervals around the mean of the estimates. Results are based on simulations of a cohort with sample size 10,000, incidence of the outcome 10%, and each scenario was simulated 10,000 times.

In Figure 4 results are presented for the NCC design. Because in the NCC design the LRM was applied as the second stage IV model, the IV estimates in the NCC design were biased irrespective of the weak or strong IVs. However, there was much more variation between estimates in the NCC design than in the cohort design (Table A2, Appendix). This variation in IV estimates between simulations decreased with increasing strength of the IV-exposure association, and also with an increasing number of controls per case (Figure 4).

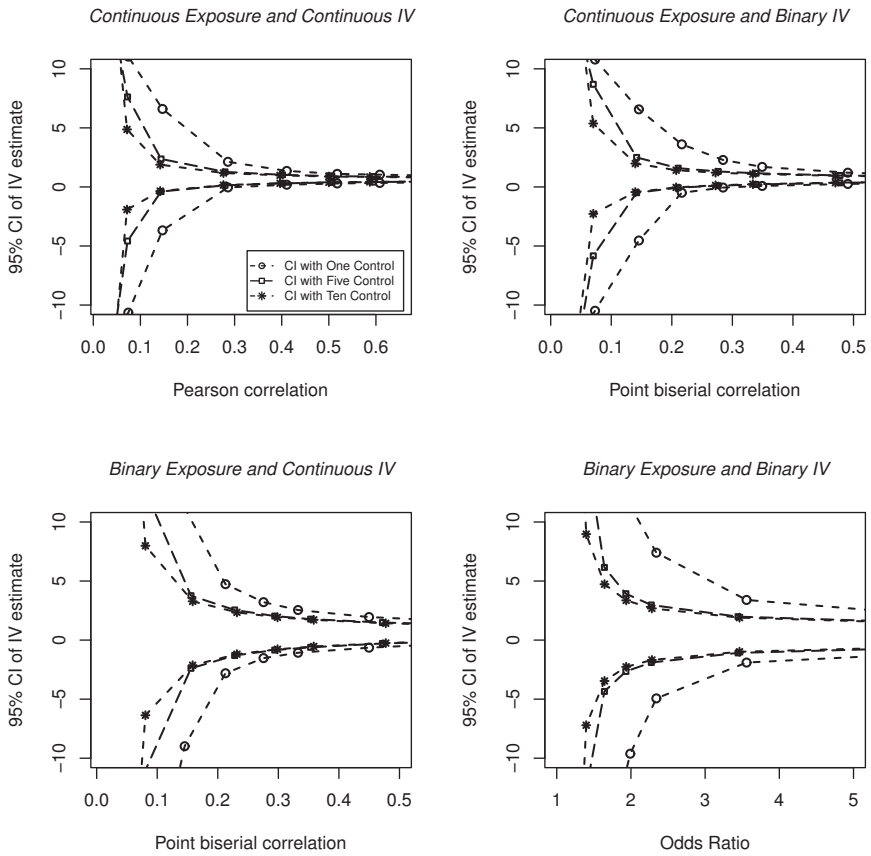


Figure 4. Variation in IV estimates in the nested case-control design

X-axis represents the association between exposure and IV. Lines indicate the 2.5 and 97.5 percentiles of the estimates in nested case-control design for different combinations of exposure and IV with case:controls ratio of 1:1, 1:5, and 1:10. Results are based on simulations of a case-control study nested within a cohort with sample size 10,000 and an incidence of the outcome of 1%. Each scenario was simulated 10,000 times.

The magnitude of bias varied for different incidences of the outcome in both designs (Table A3, Appendix for cohort design). Our simulations also revealed that the IV estimates were highly unstable when the incidence of the outcome was very rare (e.g., 1% in our simulations of a cohort of size 10,000), which was more pronounced in the NCC design than in the cohort design. Table A4 in the appendix shows the impact of sample sizes on the IV estimates in the cohort design. For each simulation setting, the RMSE decreased as sample sizes increased.

A comparison between conventional and IV analysis revealed that the IV analysis yielded unbiased results within the reasonable range of PC (> 0.15) or OR (> 2.0), which is also reflected by the lower RMSE of the IV estimates for continuous outcome and binary outcome in RD scale even if the variance of conventional estimates was lower than for IV estimates. Furthermore, the coverage rate of conventional estimates was very low due to their substantial bias, in contrast to correct coverage rates for IV estimates.

Our simulation study showed that in many situations when the IV was extremely weak (e.g., PC < 0.10 or OR < 1.5), the F-statistic value was more than 10 (a commonly used cut-off value for a weak IV), yet IV estimates were still significantly biased and highly

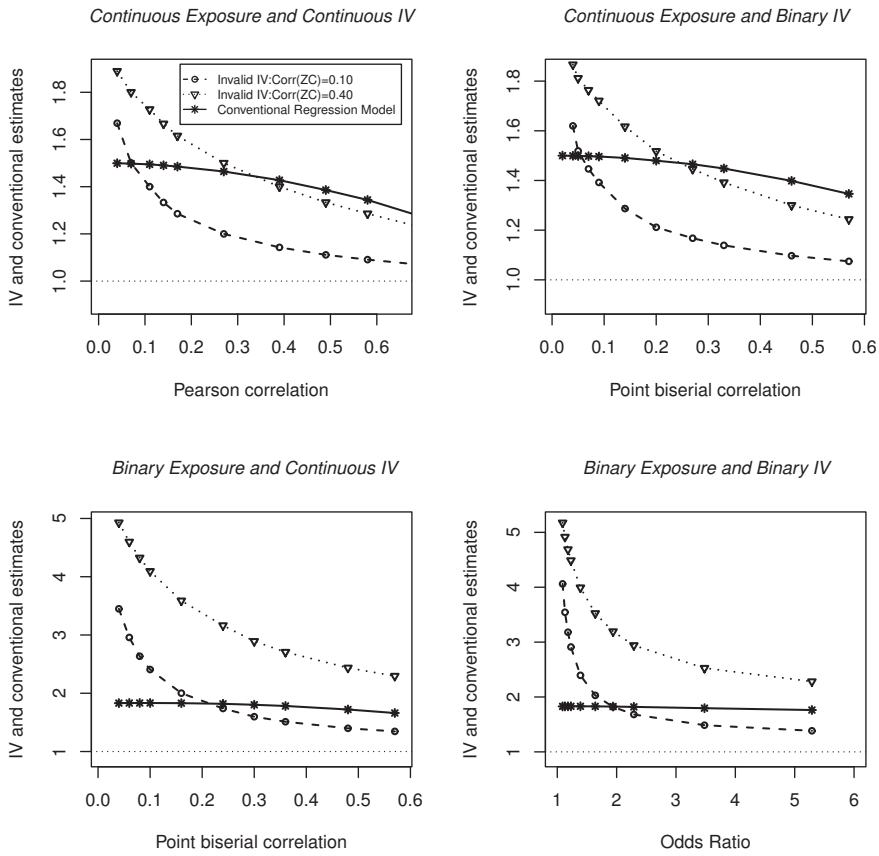


Figure 5. Mean of IV estimates and conventional estimates from simulations of a cohort design with a continuous outcome and an invalid IV (IV correlated with confounders)

X-axis represents the association between exposure and IV. The horizontal straight line (dotted line) represents the reference line (true exposure effect = 1). Results are based on simulations where the correlation between IV and confounder was p_{zc} = 0.10 and 0.40 and a sample of size 10,000 with 10,000 replications.

variable. For example, when exposure and IV were continuous and sample size 10,000, the bias was -0.070 with an F-value = 14 and *PC* = 0.04. Similarly, when both exposure and IV were binary, the bias was -0.073 with F-value = 17 and OR = 1.2. In these situations, the F-statistics values were misleading for assessing the strength of the IV as this is mostly affected by the sample sizes.

Figure 5 shows the impact of violation of the assumption that the IV is independent of confounders. It shows the patterns of bias in the cohort design for different combinations of exposure and IV with invalid IV (i.e., correlation between the IV and the confounding variable, *PC* = 0.10 and *PC* = 0.40) and estimates from the conventional regression model. In all settings, when the IV was weak and related to the confounding variable, the biases were significantly larger than in the conventional model. This pattern was more pronounced for *PC* = 0.40. However, for a stronger IV with *PC* = 0.10, the bias was lower than conventional estimates.

In all simulation settings, the pattern of bias was similar for different confounding effects, but the magnitude of bias increased with increasing the effects of confounder (values of β_c) on the exposure and outcome. Since these results are expected, we do not show these results in detail; all presented results for $\beta_c = 1$ and $\beta_x = 0.005$ (RD models).

DISCUSSION

Our simulation study shows that, in all binary/continuous combinations of IV, exposure, and outcome in the cohort and nested case-control (NCC) designs, the validity of IV analysis strongly depends on the strength of the association between IV and exposure. IV estimates are unbiased in case of a strong IV and a continuous outcome whereas conventional estimates can be extremely biased (due to unmeasured confounding). For binary outcomes, the IV estimates on the RD scale are also unbiased (again the conventional estimates are biased) but on the OR scale that are systematically biased even with strong IVs. In all cases, the variance of estimates is lower for the conventional estimates compared to IV estimates, which is partly due to bias-variance trade-off.[5] In addition, for weak IVs, large variation between IV estimates from different studies was observed.[8] In fact, the extreme variability of the estimates for very weak IVs makes it difficult to assess the accuracy of the point estimates and makes these estimates practically useless. This pattern was observed for different types (binary/continuous) of IV, exposure, and outcome in the cohort as well as the NCC design. Because of the effective sample size in the NCC design is smaller than in a cohort design, this behaviour of IV analysis was more pronounced in the NCC design, but could partly be remedied by increasing the number of controls per case.

Although the set-up of our simulations was such that the IV was independent of confounders and the IV had no direct effect on the outcome, in finite-samples random variation may cause the actual data to deviate from these assumptions, which can result in some amount of bias (conditional on the observed data) for weak IVs.[11,13,18,43] Furthermore, in

finite-samples, even a small bias of IV estimates can be amplified considerably by a weak IV, resulting in large variation between IV estimates.[6,7,12] Another aspect of our simulations that deserves attention is that the bias in the conventional estimates became smaller with increasing strength of the IV-exposure association. It is inherent to our set-up that there is an inverse relation between the strength of the IV and the magnitude of the unmeasured confounding. For details we refer to the work by Martens et al.[11]

Variability of IV estimates not only depends on the strength of the IV-exposure association, but also on the outcome. We observed that if the outcome is rare (e.g. 1% in a cohort of 10,000), IV estimates are highly variable. This is even more pronounced in the NCC design due to relatively small effective sample size. Additionally, the instability of the IV estimates could also be increased due to the two-step analysis: variability in the first-stage model can result in even more variation in the second modelling step.[22] This problem can be reduced by increasing the sample size, the number of cases, and increasing the number of controls in the NCC design, as long as the IV-exposure association is sufficiently strong. These findings suggest that IV analysis should be carefully performed in the case of rare outcomes especially in the NCC design.

Our NCC design findings suggest that selection and construction of instrumental variables in the NCC study is in line with the cohort study if the control subjects are sampled appropriately. However, as selection of appropriate controls is a major challenge in practical settings, cautions should be taken when selecting the controls. Our findings also suggest increasing the number of controls in order to increase effective sample size and reduce variability of the IV estimates.

In simulations with a binary outcome, we used logistic and linear regression models in the second-stage model, and observed that the IV estimates on the OR scale are biased even with a strong IV. Others also stated that IV estimates are biased with non-linear models (e.g. logistic regression).[44–47] Reasons for this include non-collapsibility of the odds ratio,[39,48] model misspecification,[19] and mean and variance dependency of the logistic regression model.[47] We compared the IV estimates on the OR scale with a (reference) marginal exposure effect instead of the conditional exposure effect; still, IV estimates were different from MSM estimates, which might be due to the fact that MSMs estimates are sensitive to weights especially for continuous exposure[49] or due to the random variation or model misspecification. However, the estimates on the RD scale are unbiased when IV-exposure association is strong enough. This finding is in line with previous literature.[5] Obviously, this result strongly relies on the fact that the linear model was used for generating and simulating the data (i.e., no model misspecification) and generalization of this result is conditional on appropriate specification of the second-stage model in empirical data.

Other studies have previously shown that if the F-statistic value of the first-stage regression model is 10, the IV is strong enough and the bias of IV estimates is negligible.[11,16,26,50,51] However, in our simulations the IV estimates were still biased and estimates were

extremely variable when the values of the F-statistic were around 10.[8] This was observed in all combinations of the exposure and IV, suggesting that a cut-off value for the F-statistic value around 10 is inadequate to identify a weak IV. Hence, only the F-statistic should not be considered as a reliable criterion for assessing the strength of the IV. We recommend researchers to report both the F-statistic value and the association between exposure and IV, by means of the Pearson's correlation, point bi-serial correlation or odds ratio.[8,10] We stress however, that interpretation of these measures of association depends on the effective sample size.

We also considered a situation where the IV is invalid (i.e., correlated with the confounding variable). We found that the amount of bias can be substantial even with a strong IV if the IV is associated with unmeasured confounders. Additionally, for a weak IV with a small association with a confounder, the bias is significantly higher but for a strong IV, the bias is smaller than the bias of a conventional model. Although in the latter situation the IV estimates are less biased, a small association between IV and confounder may lead to the violation of an assumption that the IV has no direct effect on the outcome. In that case, there is no guarantee that the IV analysis consistently estimates the average effect of exposure on outcome.[12] Martens et al[11] and Rassen et al[14] already reported that violation of any IV assumption can magnify the bias due to confounding. The results of our simulations confirm that an IV analysis is not a valid analysis (and estimated exposure effects are biased) if the IV is associated with confounders (distribution of confounders does not balance between IV levels) even when the IV is strongly related to exposure.

We did not assess the bias and variability of IV estimates with heterogeneous exposure effects as well as time-varying IV, exposure, and confounder. Therefore, it may be interesting to assess the trend of bias and variability due to a weak IV in these settings in future research.

In conclusion, for different binary/continuous combinations of IV, exposure, and outcome in the cohort and NCC designs, IV estimates performed uniformly poor in case of weak IV-exposure associations. When the IV is valid and has a strong association with exposure, IV methods provide unbiased exposure effect (for continuous outcome as well as binary outcome on a RD scale) whereas conventional analysis is substantially biased due to unmeasured confounding. For weaker IVs, estimates become highly variable. This variability was more pronounced for rare outcomes, especially in the NCC design. To some extent, this can be remedied by increasing the sample size, or by increasing the number of controls per case in a case-control study. Since the effective sample size in the NCC design is smaller than in the cohort design, in order to achieve stable estimates, the association between exposure and IV should even be stronger in NCC studies than in cohort studies. When the IV is not independent of confounders, this may result in severe bias even with strong IVs. We recommend researchers to routinely evaluate and report the strength of the IV in order to evaluate the potential for bias of IV estimates due to a weak IV.

REFERENCES

1. 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; 17(3): 268-75.
2. Wang PS, Schneeweiss S, Avorn J, Fischer MA, Mogun H, Solomon DH, Brookhart MA. Risk of death in elderly users of conventional vs. atypical antipsychotic medications. *N Engl J Med* 2005; 353(22): 2335-41.
3. 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; 54(11): 3390-8.
4. 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. *Can Med Assoc J* 2007; 176(5): 627.
5. Ionescu-Ittu R, Delaney JAC, Abrahamowicz M. Bias–variance trade-off in pharmacoepidemiological studies using physician-preference-based instrumental variables: a simulation study. *Pharmacoepidemiol Drug Saf* 2009; 18(7): 562-71.
6. Shetty KD, Vogt WB, Bhattacharya J. Hormone replacement therapy and cardiovascular health in the United States. *Med Care* 2009; 47(5): 600.
7. Suh HS, Hay JW, Johnson KA, Doctor JN. Comparative effectiveness of statin plus fibrate combination therapy and statin monotherapy in patients with type 2 diabetes: use of propensity-score and instrumental variable methods to adjust for treatment-selection bias. *Pharmacoepidemiol Drug Saf* 2012; 21(5): 470-84.
8. Ionescu-Ittu R, Abrahamowicz M, Pilote L. Treatment effect estimates varied depending on the definition of the provider prescribing preference-based instrumental variables. *J Clin Epidemiol* 2012; 65(2): 155-62.
9. 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; 45(10): S116-22.
10. Pratt N, Roughead EE, Ryan P, Salter A. Antipsychotics and the risk of death in the elderly: An instrumental variable analysis using two preference based instruments. *Pharmacoepidemiol Drug Saf* 2010; 19(7): 699-707.
11. Martens EP, Pestman WR, De Boer A, Belitser SV, Klungel OH. Instrumental variables: Application and limitations. *Epidemiology* 2006; 17(3): 260-7.
12. Hernán MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? *Epidemiology* 2006; 17(4): 360-72.
13. Angrist J, Imbens G, Rubin DB. Identification of causal effects using instrumental variables *J Am Stat Assoc* 1996; 91(434): 444-72.
14. Rassen JA, Brookhart MA, Glynn RJ, Mittleman MA, Schneeweiss S. Instrumental variables I: instrumental variables exploit natural variation in nonexperimental data to estimate causal relationships. *J Clin Epidemiol* 2009; 62(12): 1226-32.
15. Brookhart MA, Rassen JA, Schneeweiss S. Instrumental variable methods in comparative safety and effectiveness research. *Pharmacoepidemiol Drug Saf* 2010; 19(6): 537-44.
16. Stock JH, Wright JH, Yogo M. A survey of weak instruments and weak identification in generalized method of moments. *J Bus Econ Stat* 2002; 20(4): 518-29.
17. Zivot E, Startz R, Nelson CR. Valid confidence intervals and inference in the presence of weak instruments. *International Economic Review* 1998; 39(4): 1119-44.

18. Bound J, Jaeger DA, Baker RM. Problems with instrumental variables estimation when the correlation between the instruments and the endogeneous explanatory variable is weak. *J Am Stat Assoc* 1995; 90(430): 443-50.
19. Angrist JD, Krueger AB. Instrumental variables and the search for identification: From supply and demand to natural experiments. *J Econ Perspect* 2001; 15(4): 69-85.
20. Chao J, Swanson NR. Alternative approximations of the bias and MSE of the IV estimator under weak identification with an application to bias correction. *J Econ* 2007; 137(2): 515-55.
21. Stock JH, Yogo M. Testing for weak instruments in linear IV regression. NBER Technical Working Paper No. 284, Issued in November 2002, http://www.nber.org/papers/t0284.pdf (date of access 23 July 2014).
22. Rassen JA, Brookhart MA, Glynn RJ, Mittleman MA, Schneeweiss S. Instrumental variables II: instrumental variable application-in 25 variations, the physician prescribing preference generally was strong and reduced covariate imbalance. *J Clin Epidemiol* 2009; 62(12): 1233-41.
23. Pierce BL, Ahsan H, Vanderweele TJ. Power and instrument strength requirements for Mendelian randomization studies using multiple genetic variants. *Int J Epidemiol* 2011; 40(3): 740-52.
24. Crown WH, Henk HJ, Vanness DJ. Some cautions on the use of instrumental variables estimators in outcomes research: How bias in instrumental variables estimators is affected by instrument strength, instrument contamination, and sample size. *Value in Health* 2011; 14(8): 1078-84.
25. Chiba Y. Bias analysis of the instrumental variable estimator as an estimator of the average causal effect. *Contemporary Clinical Trials* 2010; 31(1): 12-7.
26. Staiger D, Stock JH. Instrumental variables regression with weak instruments. *Econometrica* 1997; 65(3): 557-86.
27. Andrews DWK, Stock JH. Inference with weak instruments. *NBER Technical Working Paper No. 313*, 2005.
28. Abrahamowicz M, Beauchamp M-, Ionescu-Ittu R, Delaney JAC, Pilote L. Reducing the variance of the prescribing preference-based instrumental variable estimates of the treatment effect. *Am J Epidemiol* 2011; 174(4): 494-502.
29. R Development Core Team. R: A language and environment for statistical computing. *R Foundation for Statistical Computing: Vienna, Austria* 2010; ISBN 3-900051-07-0(01/19).
30. Burton A, Altman DG, Royston P, Holder RL. The design of simulation studies in medical statistics. *Stat Med* 2006; 25(24): 4279-92.
31. Davies NM, Smith GD, Windmeijer F, Martin RM. Issues in the reporting and conduct of instrumental variable studies: A systematic review. *Epidemiology* 2013; 24(3): 363-9.
32. Chen Y, Briesacher BA. Use of instrumental variable in prescription drug research with observational data: a systematic review. *J Clin Epidemiol* 2011; 64(6): 687-700.
33. Cole JA, Norman H, Weatherby LB, Walker AM. Drug copayment and adherence in chronic heart failure: Effect on cost and outcomes. *Pharmacotherapy* 2006; 26(8 I): 1157-64.
34. Velentgas P, Dreyer NA, Nourjah P, Smith SR, Torchia MM, eds. Developing a Protocol for Observational Comparative Effectiveness Research: A User's Guide. AHRQ Publication No. 12(13)-EHC099. Rockville, MD: Agency for Healthcare Research and Quality; January 2013. www.effectivehealthcare.ahrq.gov/Methods-OCER.cfm.
35. Angrist JD. Estimation of limited dependent variable models with dummy endogenous regressors. *J Bus Econ Stat* 2001; 19(1): 2-28.
36. Palmer TM, Sterne JAC, Harbord RM, et al. Instrumental variable estimation of causal risk ratios and causal odds ratios in mendelian randomization analyses. *Am J Epidemiol* 2011; 173(12): 1392-403.
37. Hennessy S, Leonard CE, Palumbo CM, Shi X, Ten Have TR. Instantaneous preference was a stronger instrumental variable than 3- and 6-month prescribing preference for NSAIDs. *J Clin Epidemiol* 2008; 61(12): 1285-8.

38. Klugh HE. Statistics: The essentials for research. Lawrence Erlbaum; 1986.
39. Greenland S, Robins JM, Pearl J. Confounding and collapsibility in causal inference. *Stat Sci* 1999; : 29-46.
40. Robins JM, Hernán MÁ, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology* 2000; 11(5): 550-60.
41. Fang G, Brooks JM, Chrischilles EA. Apples and oranges? Interpretations of risk adjustment and instrumental variable estimates of intended treatment effects using observational data. *Am J Epidemiol* 2011; 175(1): 60-5. doi:10.1093/aje/kwr283.
42. Xiao Y, Moodie EE, Abrahamowicz M. Comparison of Approaches to Weight Truncation for Marginal Structural Cox Models. *Epidemiological Methods* 2013; : 1-20.
43. Murray MP. Avoiding invalid instruments and coping with weak instruments. *The journal of economic perspectives* 2006; 20(4): 111-32.
44. Cai B, Small DS, Have TRT. Two-stage instrumental variable methods for estimating the causal odds ratio: Analysis of bias. *Stat Med* 2011; 30(15): 1809-24.
45. Henneman TA, Van Der Laan MJ, Hubbard AE. Estimating causal parameters in marginal structural models with unmeasured confounders using instrumental variables. *(U.C. Berkeley Division of Biostatistics Working Paper Series, paper 104). Berkeley, CA: The Berkeley Electronic Press* 2002.
46. Palmer TM, Thompson JR, Tobin MD, Sheehan NA, Burton PR. Adjusting for bias and unmeasured confounding in Mendelian randomization studies with binary responses. *Int J Epidemiol* 2008; 37(5): 1161-8.
47. Didelez V, Meng S, Sheehan N. On the bias of IV estimators for Mendelian Randomisation. 2008; *submitted manuscript, http://137.222.80.3/research/stats/reports/2008/0820.pdf*.
48. Burgess S. Identifying the odds ratio estimated by a two-stage instrumental variable analysis with a logistic regression model. *Stat Med* 2013; 32(27): 4726-47.
49. Daniel RM, Cousens SN, De Stavola BL, Kenward MG, Sterne JAC. Methods for dealing with time-dependent confounding. *Stat Med* 2012.
50. Han C, Schmidt P. The asymptotic distribution of the instrumental variable estimators when the instruments are not correlated with the regressors. *Economics Letters* 2001; 74(1): 61-6.
51. Angrist JD, Pischke JS. Mostly harmless econometrics: An empiricist's companion. Princeton Univ Pr; 2008.

APPENDIX

Table A1. Estimates and RMSE of simulation of a binary outcome in the cohort design (IV estimates, MSMs, and Conventional Estimates)

IV continuous X continuous Y binary							IV binary X continuous Y binary						
	IV Method ¥		MSMs**		Conventional Method*			IV Method		MSMs**		Conventional Method	
PC	Estimate	RMSE	Estimate	RMSE	Estimate	RMSE	PBC	Estimate	RMSE	Estimate	RMSE	Estimate	RMSE
0.14	0.515	0.237	0.772	0.097	1.110	0.419	0.07	0.504	0.384	0.777	0.103	1.119	0.427
0.27	0.515	0.195	0.756	0.083	1.082	0.390	0.14	0.513	0.242	0.772	0.098	1.111	0.419
0.39	0.514	0.187	0.733	0.064	1.040	0.348	0.20	0.513	0.209	0.765	0.091	1.098	0.406
0.49	0.513	0.185	0.711	0.048	0.992	0.300	0.27	0.512	0.197	0.755	0.082	1.080	0.388
0.58	0.510	0.186	0.691	0.040	0.944	0.252	0.33	0.514	0.190	0.749	0.075	1.060	0.368
IV continuous X binary Y binary							**IV binary X binary Y binary**						
	IV Method		MSMs		Conventional Method			IV Method		MSMs		Conventional Method	
PBC	Estimate	RMSE	Estimate	RMSE	Estimate	RMSE	OR	Estimate	RMSE	Estimate	RMSE	Estimate	RMSE
0.08	0.541	0.827	0.636	0.107	1.410	0.721	2.29	0.527	0.369	0.640	0.105	1.388	0.700
0.16	0.550	0.431	0.637	0.106	1.396	0.707	2.71	0.549	0.312	0.641	0.101	1.371	0.685
0.24	0.555	0.310	0.639	0.104	1.374	0.686	3.20	0.534	0.286	0.643	0.100	1.362	0.673
0.30	0.559	0.253	0.639	0.102	1.347	0.658	3.48	0.528	0.277	0.645	0.100	1.355	0.667
0.36	0.563	0.221	0.639	0.100	1.316	0.628	5.30	0.523	0.242	0.650	0.095	1.314	0.626

IV: instrumental variable; X: exposure; Y: outcome; PC: Pearson's correlation; PBC: point bi-serial correlation; OR: odds ratio; RMSE: root mean square error, MSMs: marginal structural models; sample size: n = 10000 and number of simulations = 10000; ¥IV estimate from two-stage IV models and the IV is independent of the confounders (valid IV); *Conventional method: logistic regression model; True exposure effect = log(2). **Truncation of weights at 0.5 and 99.5 percentiles.

Table A2. Comparison of bias and standard deviation of IV estimates based on full cohort versus NCC design (with various case:control ratios)

	IV continuous X continuous Y binary						IV binary X continuous Y binary						
			NCC							NCC			
	Cohort design		case:control=1:1		case:control=1:10		Cohort design			case:control=1:1		case:control=1:10	
PC	Estimate	SD	Estimate	SD	Estimate	SD	PBC	Estimate	SD	Estimate	SD	Estimate	SD
0.14	0.631	0.471	1.006	20.814	0.669	0.568	0.07	0.610	1.033	0.377	62.330	0.807	6.674
0.28	0.626	0.234	0.720	3.391	0.643	0.262	0.14	0.627	0.498	0.877	22.395	0.673	0.604
0.40	0.619	0.153	0.666	0.305	0.635	0.171	0.21	0.632	0.330	0.694	17.919	0.652	0.367
0.50	0.623	0.120	0.646	0.208	0.631	0.125	0.27	0.629	0.244	0.799	4.331	0.647	0.273
0.59	0.619	0.095	0.647	0.179	0.632	0.108	0.33	0.631	0.198	0.732	3.549	0.647	0.220
	IV continuous X binary Y binary						IV binary X binary Y binary						
			NCC		NCC					NCC		NCC	
	Cohort design		case:control=1:1		case:control=1:10		Cohort design			case:control=1:1		case:control=1:10	
PBC	Estimate	SD	Estimate	SD	Estimate	SD	OR	Estimate	SD	Estimate	SD	Estimate	SD
0.08	0.492	2.399	1.595	418.358	0.890	19.715	2.28	0.530	0.330	0.541	27.864	0.474	1.100
0.16	0.507	1.199	1.794	95.305	0.545	1.365	2.71	0.549	0.312	0.547	8.794	0.477	0.909
0.23	0.518	0.819	0.662	25.933	0.543	0.902	3.20	0.535	0.286	0.419	5.785	0.444	0.781
0.30	0.529	0.635	0.648	2.073	0.546	0.691	3.46	0.530	0.223	0.470	5.267	0.438	0.731
0.36	0.537	0.529	0.588	1.359	0.552	0.573	5.32	0.525	0.172	0.458	1.310	0.414	0.571

IV: instrumental variable; X: exposure; Y: outcome; PC: Pearson's correlation; PBC: point bi-serial correlation; OR: odds ratio; NCC: Nested case-control design; SD: standard deviation (variability between estimates from 10,000 simulation runs). IV estimate from two-stage IV models and the IV is independent of the confounders (valid IV); True exposure effect = log(2). Incidence of the outcome is 1%.

Table A3. Impact of incidence of the outcome on the IV estimates in a cohort design with continuous outcome

IV continuous X continuous Y binary							IV binary X continuous Y binary						
	Incidence=1%		Incidence=10%		Incidence=25%			Incidence=1%		Incidence=10%		Incidence=25%	
PC	Estimate	RMSE	Estimate	RMSE	Estimate	RMSE	PBC	Estimate	RMSE	Estimate	RMSE	Estimate	RMSE
0.14	0.631	0.475	0.515	0.237	0.463	0.253	0.07	0.610	1.036	0.503	0.377	0.453	0.325
0.28	0.626	0.244	0.515	0.217	0.465	0.235	0.14	0.627	0.503	0.508	0.316	0.460	0.256
0.40	0.619	0.170	0.515	0.194	0.465	0.231	0.20	0.632	0.335	0.513	0.240	0.462	0.242
0.50	0.623	0.139	0.512	0.188	0.465	0.229	0.27	0.629	0.253	0.514	0.207	0.462	0.237
0.59	0.619	0.121	0.513	0.185	0.466	0.229	0.33	0.631	0.207	0.515	0.195	0.463	0.234
IV continuous X binary Y binary							IV binary X binary Y binary						
	Incidence=1%		Incidence=10%		Incidence=25%			Incidence=1%		Incidence=10%		Incidence=25%	
PBC	Estimate	RMSE	Estimate	RMSE	Estimate	RMSE	OR	Estimate	RMSE	Estimate	RMSE	Estimate	RMSE
0.08	0.492	2.408	0.540	0.833	0.582	0.576	2.29	0.482	1.014	0.532	0.365	0.574	0.248
0.16	0.507	1.213	0.543	0.669	0.584	0.304	2.71	0.509	0.853	0.528	0.330	0.565	0.223
0.24	0.518	0.838	0.550	0.434	0.584	0.222	3.20	0.494	0.743	0.526	0.295	0.577	0.205
0.30	0.529	0.656	0.554	0.314	0.584	0.187	3.48	0.487	0.701	0.531	0.275	0.565	0.200
0.36	0.537	0.552	0.558	0.257	0.584	0.168	5.30	0.487	0.555	0.526	0.239	0.556	0.182

IV: instrumental variable; X: exposure; Y: outcome; PC: Pearson's correlation; PBC: point bi-serial correlation; OR: odds ratio; RMSE: root mean square error, number of simulations = 10000. IV estimate from two-stage IV models and the IV is independent of the confounders (valid IV); True exposure effect = log(2).

Table A4. Impact of sample size on the IV estimates in a cohort design with continuous outcome

IV continuous X continuous Y continuous							IV binary X continuous Y continuous						
	n=1000		n=5000		n=10000			n=1000		n=5000		n=10000	
PC	Estimate	RMSE	Estimate	RMSE	Estimate	RMSE	PBC	Estimate	RMSE	Estimate	RMSE	Estimate	RMSE
0.14	0.970	0.265	0.994	0.103	0.996	0.072	0.07	0.679	21.446	0.975	0.229	0.988	0.153
0.27	0.994	0.116	0.998	0.051	1.000	0.036	0.14	0.971	0.268	0.994	0.105	0.997	0.073
0.39	0.997	0.076	0.999	0.034	0.999	0.024	0.20	0.989	0.160	0.997	0.069	0.999	0.048
0.49	0.999	0.057	1.000	0.025	1.000	0.018	0.27	0.994	0.117	0.999	0.051	0.999	0.036
0.58	0.999	0.045	1.000	0.020	1.000	0.014	0.33	0.997	0.093	0.999	0.041	1.000	0.029
IV continuous X binary Y continuous							IV binary X binary Y continuous						
	n=1000		n=5000		n=10000			n=1000		n=5000		n=10000	
PBC	Estimate	RMSE	Estimate	RMSE	Estimate	RMSE	OR	Estimate	RMSE	Estimate	RMSE	Estimate	RMSE
0.08	0.594	22.619	0.967	0.519	0.987	0.353	2.29	0.983	0.467	0.995	0.203	0.999	0.141
0.16	0.974	0.602	0.991	0.252	0.997	0.176	2.71	0.989	0.386	0.996	0.170	1.000	0.119
0.24	0.992	0.388	0.995	0.171	0.999	0.120	3.20	0.993	0.332	0.997	0.148	1.000	0.103
0.30	0.997	0.298	0.997	0.133	0.999	0.093	3.48	0.994	0.307	0.997	0.137	1.000	0.096
0.36	0.999	0.247	0.998	0.111	1.000	0.078	5.30	0.998	0.235	0.998	0.105	1.000	0.074

IV: instrumental variable; X: exposure; Y: outcome; PC: Pearson's correlation; PBC: point bi-serial correlation; OR: odds ratio; RMSE: root mean square error, number of simulations = 10000. IV estimates from two-stage IV models and the IV is independent of the confounders (valid IV); True exposure effect = 1.

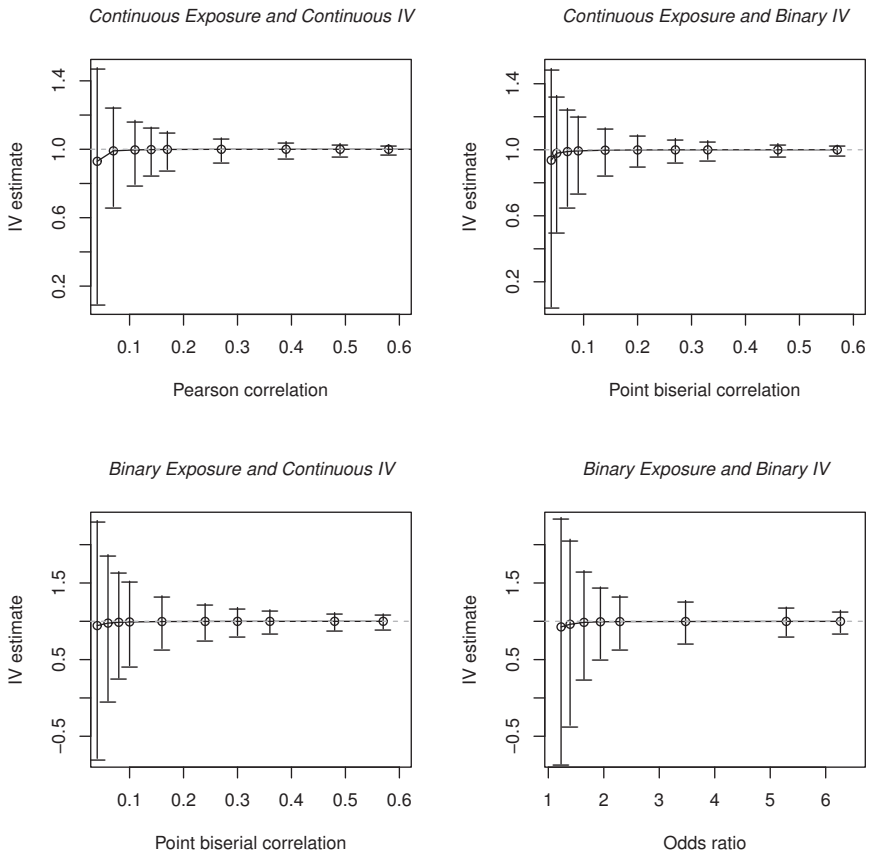


Figure A1. Mean and variation of IV estimates from simulations of a cohort design with a continuous outcome and different combinations of exposure and IV

X-axis represents the association between exposure and IV. The horizontal straight line (dotted line) represents the reference line (true exposure effect = 1). Vertical bars indicate the range for 95% of the estimates and are based on the 2.5 and 97.5 percentiles of the distribution of the IV estimates. Results are based on simulations of a cohort with sample size 10,000, and each scenario was simulated 10,000 times.

CHAPTER 3.2

Quantitative Falsification of Instrumental Variables Assumption Using Balance Measures

Md Jamal Uddin*, M Sanni Ali*, Rolf HH Groenwold, Wiebe R Pestman, Svetlana V Belitser, Arno W Hoes, Anthonius de Boer, Kit CB Roes, Olaf H Klungel,

*Both authors contributed equally to this work

The modified version of this chapter has been published as a research letter in Epidemiology, 2014; 25(5):770-72.

ABSTRACT

BACKGROUND: Instrumental variable (IV) analysis can in theory control for unmeasured confounding in non-randomized studies. We aimed to explore the usefulness of balance measures to quantitatively falsify one of the assumptions that the IV is independent of confounders.

METHODS: We conducted a simulation study to assess the performance of balance measures commonly used in propensity score methods (i.e., standardized difference) to quantitatively falsify this assumption. We simulated cohorts of varying sample sizes, binary IV and exposure, continuous outcome, and several confounders. Data were analysed using the two-stage least squares method. The balance of confounders across IV levels was assessed using the standardized difference.

RESULTS: Bias of IV estimates increased with weaker IVs (i.e., weaker association between IV and exposure) and increasing values of the standardized difference (i.e., decreasing balance of confounders across IV levels). IV estimates were more biased than conventional regression estimates with increasing values of the standardized difference, and a weak IV amplified this bias.

CONCLUSIONS: Balance measures that are commonly used in propensity score methods can be useful tools to falsify an important assumption underlying IV analysis, i.e., that the IV is independent of confounders. However, these balance measures only quantify the balance of measured confounders and researchers should complement it with theoretical justifications for balance on unmeasured confounders. If measured confounders are imbalanced between IV categories, such imbalance is likely to exist in unmeasured confounders as well and adjusting for measured confounders in the IV models may result in more biased estimate compared to conventional regression estimate; hence, researchers should consider refraining from IV analysis.

INTRODUCTION

Instrumental variable (IV) analysis might be an attractive method to adjust for confounding in non-randomized studies, since it potentially controls for both measured and unmeasured confounding.[1,2] An IV is a variable that is associated with the exposure under study and related to the outcome only through the exposure.[1,2] This implies that an IV should satisfy three basic assumptions: i) the IV is associated with the exposure under study; ii) the IV affects the outcome only through the exposure;[2–4] and iii) the IV is independent of confounders.[2,3] If these assumptions are satisfied, with additional assumption (homogeneous treatment effect, see later) IV analysis may provide consistent estimate of exposure effect on the outcome.[2] However, if one of the basic assumptions is violated, the IV estimate can be severely biased and inconsistent.[1,2]

To check the first assumption, statistical tools such as the F-statistic,[1,5–7] R squared,[8] pseudo-R-squared, [9] and the odds ratio[9,10] have been used. There is no well-established method for checking the second and third assumptions and several authors[1,9,11,12] argued that these assumptions are unverifiable or directly untestable from data as they involve unmeasured confounders.[2,13] On the other hand, Glymour et al.[14] suggested several approaches (e.g., leverage prior causal assumptions) for evaluating the validity of IV, although, in certain situations, they might fail to identify a biased IV or inappropriately suggest that a valid IV is biased. Moreover, several other authors[15–25] provided supportive evidence for the third assumption, by describing the balance of measured patient characteristics between IV categories. Alternatively, an imbalance in measured patient characteristics can falsify the third assumption and help researchers assess whether it is appropriate to proceed with IV analysis.[26]

In this article, we propose familiar and easy to apply methods that will help researchers to falsify the third assumption by assessing the independence between the IV and measured confounders. These methods are based on balance measures commonly used in propensity score (PS) methods.[27–29] In PS methods,[30] balance measures such as the standardized difference,[27–29] the Kolmogorov-Smirnov distance, and the overlapping coefficient[28,29] can be used to quantify balance of measured confounder distributions between exposure groups. These balance measures are chosen since they are robust with respect to sample size and known among epidemiologists.[28,29,31] If measured confounders are insufficiently balanced between the IV categories, the IV and measured confounders are not independent, which means that the third assumption is violated; hence, IV analysis is not appropriate. However, if measured confounders are balanced, investigators should rely on substantial background knowledge to argue that such balance could be carried over to unmeasured confounders although it cannot be verified from the data.[15,25] The objective of this study was to explore the usefulness of balance measures to quantitatively falsify the assumption that the IV should be independent of confounders. In addition, we illustrated this method using an empirical example on the relation between inhaled long-acting beta$_2$ adrenoceptor agonists and the risk for acute myocardial infarction.

METHODS

Balance measures for measured confounders

Balance measures have been used in PS methods to assess balance of confounders between treatment groups.[28,29] For a detailed explanation of the balance measures, we refer to the literature.[27–29] We used the standardized difference (SDif), the Kolmogorov-Smirnov (KS) distance, and the overlapping coefficient (OVL) as balance measures, but we only report results for the SDif owing to its common application in the medical literature, better performance in various scenarios (e.g., covariate distribution and sample size), and simplicity of calculation compared to other balance measures.[32,33] For binary IVs and binary confounders, SDif is the difference in proportions of the confounder between IV categories standardized to the variation in the confounding variable (i.e. the standard deviation). For

binary IVs and continuous confounders, SDif is the difference in the means of confounders standardized to the variation. SDif has a minimum value of zero ('perfect' balance) but no maximum value. An imbalance in measured confounders between IV categories indicated by the balance measure (e.g. SDif > 0.10, an arbitrary cut-off)[34] means that the third assumption is violated.

Simulation setting

We used Monte Carlo simulations to assess the third assumption of IV analysis using the SDif. The scenarios we considered included binary IV, binary exposure, continuous outcome, and continuous confounders. We used the following notations: Y denotes the outcome, X denotes the exposure, Z denotes the IV, and C and U denote set of measured and unmeasured confounding variables, respectively. We used statistical software R (Windows, version 2.15.1) for simulations and analyses.[35]

Data generation

First, we generated four continuous confounders (C_1, C_2, C_3, and C_4) using the multivariate normal distribution (MVND) with mean 0 and variance 1. The correlation coefficient between confounders was varied between 0 and 0.4. A binary IV was generated based on the following logistic models (equation 1).

$$\text{logit}[\text{Prob}(Z=1|C)] = \alpha_0 + \alpha_1 C_{1i} + \alpha_2 C_{2i} + \alpha_3 C_{3i} + \alpha_4 C_{4i} \quad [1]$$

The values for α_1- α_4 were varied between 0.0 and 0.60 to induce different association between IV and confounders. α_0 was set to -0.42 in the logistic model (equation 1) in order to achieve 40% prevalence of the binary IV. Next, a binary exposure was generated based on logistic model (equation 2).

$$\text{logit}[\text{Prob}(X=1|Z,C)] = \beta_0 + \beta_z Z_i + \beta_1 C_{1i} + \beta_2 C_{2i} + \beta_3 C_{3i} + \beta_4 C_{4i} \quad [2]$$

The value of β_0 was set to -1.5 so that nearly 50% of the subjects were treated, the values of β_z was varied between 0.20 to 2.50 to induce different associations between the IV and the exposure, and β_1 through β_4 were set to 1.5 in different scenarios.

A continuous outcome, Y, was generated using the following model (equation 3):

$$Y_i = \delta_0 + \delta_x X_i + \delta_1 C_{1i} + \delta_2 C_{2i} + \delta_3 C_{3i} + \delta_4 C_{4i} + \varepsilon_i; \quad \text{for } i = 1,2,.....n \quad [3]$$

where X indicates the exposure variable generated previously (equation 2) and the variable C_1, C_2, C_3, and C_4 denote the confounding factors. δ_0, δ_x, and δ_1 - δ_4 denote the intercept (set to 1.0), the true exposure effect (set to 1.0), and the effects of the confounders on the outcome (set to 1.5), respectively. ε is the error term for the outcome, which follows a normal distribution with mean zero and variance of unity.

We distinguished three scenarios which are schematically presented in Figures 1a-b, 2a-b, and 3a-b, using causal diagrams.

Scenario 1: All confounders are measured (no unmeasured confounding) (Figure 1a and 1b);

Scenario 2: One of the confounders (C_4 = U) is unmeasured; this unmeasured confounder has no association with the measured confounders (C_1, C_2, and C_3) (Figure 2a and 2b).

Scenario 3: One of the confounders (C_4 = U) is unmeasured; however, this confounder is associated with the measured confounders (C_1, C_2, and C_3) (Figure 3a and 3b).

In scenarios 2 and 3, when there is unmeasured confounding (U), the measured confounder C_4 was considered to be unmeasured in the analysis stage. Hence, no assessment was made on the balance of the distribution of this variable between IV categories. All scenarios were evaluated for a sample size of 10,000 and each scenario was replicated 10,000 times. In order to identify the average exposure effect among the study population, we assumed that the effect of exposure on the outcome was the same for all subjects.[2]

Analysis of simulated data

In all the three scenarios, the balance of measured confounders between IV groups was assessed using SDif. In the presence of unmeasured confounders, balance was only assessed on measured confounders. We analysed data using the two-stage least squares method. The first-stage model was a linear regression model, in which the exposure was the dependent and the IV was the independent variable.[36] The second-stage model was also a linear regression model, in which the outcome was regressed on the predicted exposure (i.e., the predicted value of exposure status based on equation (4)), rather than the actual exposure. These models can be summarized:

First-stage model: $\mathrm{X_i} = \gamma_0 + \gamma_1 \mathrm{Z_i} + \varepsilon_{\mathrm{i1}}$; for $i = 1,2,.....n$ [4]

Second-stage model: $Y_i = \theta_0 + \theta_x \hat{X}_i + \varepsilon_{i2}$; for $i = 1,2,.....n$ [5]

where X and Z are exposure and IV, respectively. $\hat{X}$ denotes the predicted value of the exposure, predicted from equation (4), ε_1 and ε_2 are the error terms, which follow a normal distribution with mean zero and constant variance σ^2, γ_0 and θ_0 denote intercepts in the first and second-stage models, respectively. The parameter θ_x in equation (5) is called the IV estimator, an estimate for the exposure effect on the outcome.

In IV analysis, the measured confounders can be included in both the first and the second-stage models since the conditional independence and exclusion restrictions underlying IV estimation are more likely to be valid after conditioning on covariates[6] and the precision of the estimates can be improved.[6,25,37,38] In addition, Brookhart et al.[39] suggested "reporting an unadjusted IV estimate and exploring the sensitivity of the results to the inclusion/exclusion

of covariates, particularly if there is not a strong theoretical reason to believe that they confound the instrument-outcome association". To evaluate this approach, we performed additional analyses including all measured confounders in each stage models, equations (4) and (5).

In addition, we used conventional multivariable linear regression models adjusting for all measured confounders to estimate the exposure effect on the outcome and compared the results with the IV estimates.

Bias was defined as the difference between the average of the estimated effects and the true exposure effect (i.e., 1.0). Confidence Intervals (CIs) were estimated using the standard errors of the mean of the estimates (i.e., standard deviation of the IV estimates divided by square root of the number of simulations) to identify the precision of estimating the bias of the IV estimates.

Empirical example

To illustrate the method, we used data from a pharmacoepidemiological study on the relation between inhaled long-acting beta$_2$ adrenoceptor agonists and the risk for acute myocardial infarction (AMI). For this follow up study, data from the Dutch Mondriaan database was used, which comprises general practitioners (GP) data complemented by pharmacy dispensing data and linkages to survey data on about 1.4 million patients. For this example we used GP data from adult patients with a diagnosis of asthma and/or COPD and at least one prescription of inhaled beta$_2$-agonists (long acting or short acting, LABA/SABA) or inhaled muscarinic antagonists (MA) (n = 27,459). The index date is defined for each individual patient, as the date of first prescription of an inhaled SABA/LABA or an inhaled MA after the start of valid data collection. The observation period for each patient lasts from the index date (from 1 January 2002 onwards) to the end of data collection (31 December 2009), the date of the first AMI, the date of death, whichever occurs first. A patient can switch between current, recent and past periods and between the treatment classes. A patient is a current user from the beginning of the prescription up to the calculated end date of the prescription, or a recent user during the 91 days following the calculated end date of the prescription, or a past user after 91 days following the calculated end of prescription. The choice for the 91 days in the calculation interval was based on the fact that Dutch health insurance policies cover the dispensing of the majority of drugs for three months.[40,41] The period of "past user" will expire if the patient becomes a new user or on the end of follow-up. If a patient switches between the treatment classes, a new treatment period starts at the date of the prescription of the new drug. For our analysis, we considered two groups for comparison: current LABA users and non-LABA users. Non-LABA users could be SABA/MA users or recent/past LABA users.

The outcome status (AMI) was based on GP records (the international classification of primary care, ICPC code = K75). Patients were excluded if they had any history of myocardial infarction prior to or at the start of follow up.

Co-morbidities and co-medications were assessed, from one year prior to study entry until the end of the study period, every time exposure changes (current, recent, past) and every six months (if the patient did not change his/her exposure status); hence, considered as time varying confounders. For co-morbidities, patients were classified as having the disease from the first date of diagnosis through follow up.

Physician's prescribing preference which is measured using the proportions of time for LABA prescriptions (PTLP) per GP centre was used as an IV in two-stage IV analysis.[25] The PTLP is the ratio of the follow-up time for current users of LABA under a GP to the total follow-up time under the same GP. Hence, PTLP is a continuous variable ranging between 0 and 1. Although dichotomizing a continuous variable (either IV or any other variable) is generally advised against, for illustration purpose, we dichotomized the continuous IV (PTLP). We used the median of the PTLP as a cut-off to create a binary IV. The first-stage model of the IV analysis was a linear regression model and the second-stage model was a Cox proportional hazards model.[42–44] The 95% CIs were estimated using bootstrapping (number of bootstraps=1000). The strength of the relation between IV and exposure (current LABA users versus non-LABA users) was quantified by the odds ratio. The SDif between IV categories were calculated for each potential (time-varying) confounder to assess balance of measured confounders between IV categories. In addition, crude and adjusted hazard ratios were estimated using conventional Cox proportional hazards model where confounders except gender were considered time-varying.

RESULTS

Figure 1 shows the relation between the SDif and bias of the IV estimate for the scenario without unmeasured confounding (i.e. all confounders measured). The magnitude of the bias increased when the balance on confounders between IV categories decreased (as indicated by an increase in the SDif). When IV was independent of measured confounders (indicated by intersection point in the plot that corresponds to the zero-point of the mean SDif), unadjusted IV estimates were unbiased. However, when the IV was associated with measured confounders, unadjusted IV estimates were biased even for stronger IV (e.g., α_i = 1.5 in equation 1 with the corresponding value of SDif was 0.6 and β_z = 2.5 in equation 2, the bias in the IV estimate was as high as 5.5). When IV was associated with measured confounders, the magnitude of the bias was also influenced by the strength of the IV (i.e. the association between IV and exposure, β_z). For example, for two IVs with β_z = 0.5 and 2.5, when the SDif of 0.6, the corresponding bias were close to 9 and 5.5, respectively. In the same Figure, the conventional multivariable linear regression estimates were unbiased where as those of unadjusted IV estimates were not except when IV is perfect, i.e. IV is independent of confounders corresponding to SDif of zero. However, results from the adjusted IV models (models that also included the measured confounders both in first and second stage models) were unbiased. The strength of association between the exposure and confounders influenced the magnitude of bias of IV estimate but not the balance of

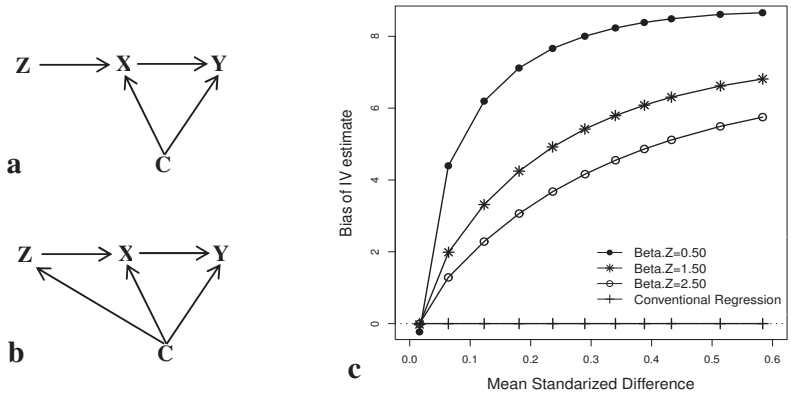


Figure 1. Directed acyclic graphs (1a and 1b, where Z = Instrumental variable, X = exposure, Y = outcome, C = measured confounder) and plot of mean standardized difference vs. bias of IV (for different strength of the IV, Beta.Z = regression coefficient of Z) and conventional regression estimates in the absence of unmeasured confounding (1c). Z is independent of C (1a) and Z is associated to C (1b). The different points on the lines indicate different correlations between IV and measured confounders (for example, at the intersection point of the three lines corresponding to the zero SDif, the IV is independent of measured confounders)

measured confounders between IV categories. We therefore only presented results for an association between the exposure and the confounders of $\beta_i = 1.5$.

Figure 2 shows the relation between the SDif and bias for the scenarios with unmeasured confounding which was independent of measured confounders. In this situation, both unadjusted IV method and the conventional multivariable regression method provided biased estimates (Figure 2c), due to association between IV and confounders (except when IV is perfect, the starting point of BetaU.Z line in the plot corresponding to the nearly zero-SDif), and the presence of unmeasured confounding, respectively. Moreover, in the case of an IV that was associated with confounders (e.g. SDif = 0.05 to 0.80), the results obtained from the conventional linear regression model were less biased than those of unadjusted IV models. Again, the magnitude of bias increased with increasing SDif. In situations where the IV was independent of the measured confounders but not of unmeasured confounders, IV estimates were still biased even though the SDif was close to zero, which is due to the fact that the SDif was determined based only on the measured confounders. When measured confounders were included as covariates in the IV models, the bias of IV estimates was close to zero when the IV was independent of the unmeasured confounders (Figure 2d). In addition, the bias for adjusted IV estimates was smaller than that of unadjusted IV estimates. Importantly, when the IV was strong but related to the confounders (both measured and unmeasured), estimates from adjusted IV models were more biased than those of conventional multivariable regression estimates.

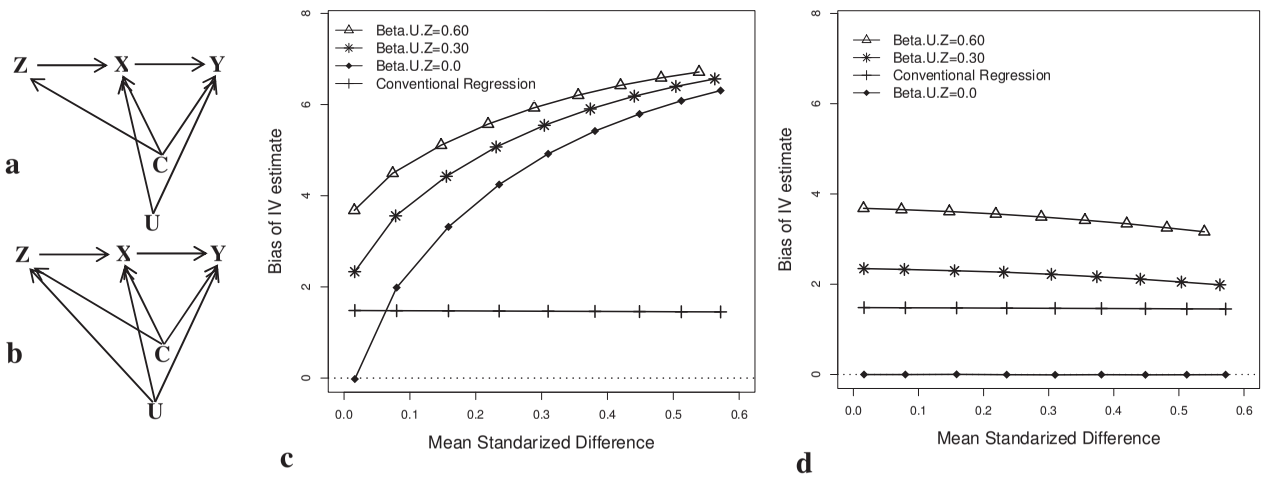


Figure 2. Directed acyclic graphs (2a and 2b, where Z = Instrumental variable, X = exposure, Y = outcome, C = measured confounder, and U = unmeasured confounder) and plots of mean standardized difference vs. bias of IV (for different association between Z and U, Beta.U.Z = regression coefficient of U) and conventional regression estimates in the presence of U that is independent of C (2c and 2d). Z is independent of U (2a) and Z is associated to U (2b). Standard IV/regression estimates (2c) and adjusted IV/regression estimates (2d).

When the unmeasured confounder was associated with measured confounders (Scenario 3, Figure 3a-b), the results showed a similar pattern as described above, with the exception that the conventional multivariable regression model was less biased (Figure 3c). Also in this scenario, the magnitude of the bias increased with increasing SDif and estimates from adjusted IV models were biased when the IV was associated with the unmeasured confounder (Figure 3d).

Empirical example: LABA use and risk of myocardial infarction

The total follow-up time was 110,146 person years and the proportion of time for LABA prescriptions per GP ranged between 0 and 0.56 (median 0.29). The mean age of patients was 52.3 (SD = 17.8) years and 447 of the patients experienced an AMI during follow-up. The odds ratio of the relation between the IV (PTLP) and exposure (current LABA use) was 1.77. The distribution of several patient characteristics differed systematically between exposure groups; hence, there was a large potential for confounding of the exposure-outcome relation (Table 1). Indeed, the crude and adjusted hazard ratios (HR) from the conventional analyses differed considerably: HR 1.39 [1.15-1.69] and 0.90 [0.73-1.10], respectively.

Some of the measured confounders were not balanced between IV categories (SDif values for age, oral corticosteroid, and disease (COPD, Asthma, or both) were 16%, 11% and 16%, respectively); hence, the third IV assumption did not seem to hold. The effect estimates from unadjusted and adjusted IV analyses were HR 1.69 [0.34-9.40] and 0.19 [0.02-1.62], respectively.

DISCUSSION

Our simulation study shows that balance measures can be used to falsify the third assumption of IV analysis, i.e., the assumption that the IV is independent of confounders. The standardized difference (SDif), a measure of the degree of balance on measured confounders between IV categories, was strongly correlated with the bias of IV estimates. Values of the SDif close to zero indicate that at least the measured confounders are balanced between IV categories. When this assumption is violated, IV analysis may result in more biased estimates than conventional regression analysis.

The magnitude of bias was associated with the strength of the IV and the balance of measured confounders between IV categories. A higher value of SDif (e.g., larger than 10%),[32] i.e., strong association between IV and measured confounders, indicates a violation of the third assumption and is associated with highly biased estimates. This bias can be remedied by including measured confounders in the IV model under the assumption of no unmeasured confounding or unmeasured confounders being independent of the IV. However, IV analysis is mainly considered in settings where unmeasured confounding cannot be ruled out. An imbalance in measured confounders as indicated by SDif would, therefore, suggest that IV analysis with or without inclusion of measured confounders would yield biased estimates. Interestingly, in the presence of unmeasured confounding that was

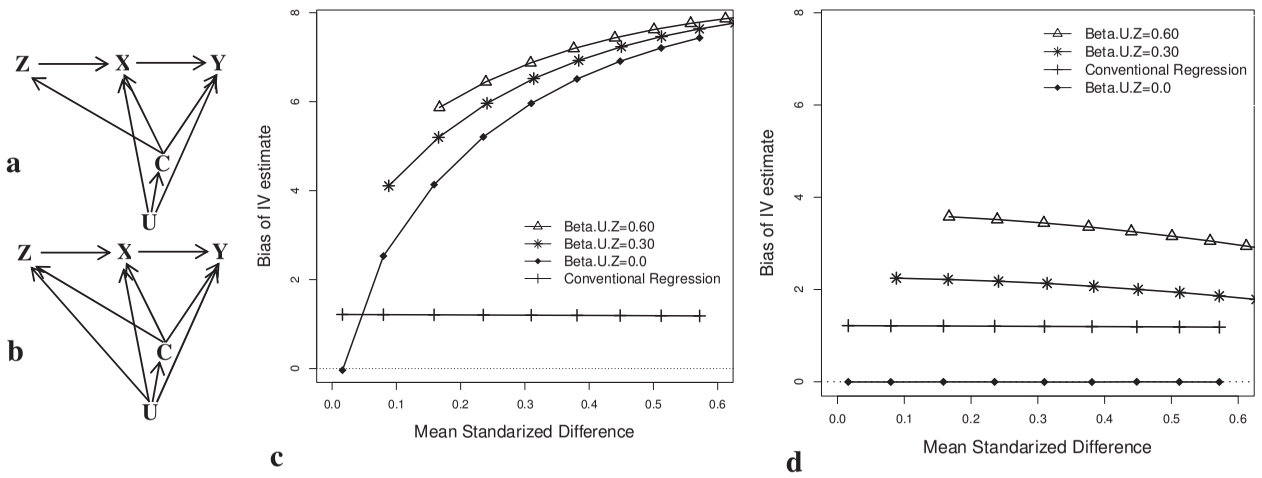


Figure 3. Directed acyclic graphs (3a and 3b, where Z = Instrumental variable, X = exposure, Y = outcome, C = measured confounder, and U = unmeasured confounder) and plots of mean standardized difference vs. bias of IV (for different association between Z and U, Beta.U.Z = regression coefficient of U) and conventional regression estimates in the presence of U that is associated with C (3c and 3d). Z is independent of U (3a) and Z is associated to U (3b) Standard IV/ regression estimates (3c) and adjusted IV/regression estimates (3d).

Table 1. Balance of confounders between exposure and instrumental variable categories

	Exposure			IV*		
	Current LABA	Non-LABA	Current LABA versus Non-LABA	PTLP = 1	PTLP = 0	PTLP*
Confounders	Mean or Frequency (%)	Mean or Frequency (%)	Standardized Difference**	Mean or Frequency (%)	Mean or Frequency (%)	Standardized Difference**
Age (Mean)	57.4	50.3	**0.414**	53.7	50.9	**0.160**
Gender (Male)	58.1	59.2	0.022	58.4	59.3	0.017
Diabetes mellitus (DM)	11.4	9.1	0.075	10.3	9.1	0.039
DM Medication	8.8	6.9	0.072	8.2	6.6	0.061
Beta-Blocker	12.6	10.3	0.072	11.9	10.0	0.060
Lipid modifying Agents	14.7	10.5	**0.128**	12.9	10.3	0.082
Stroke	1.9	1.6	0.025	1.7	1.7	0.001
Ischaemic heart disease	6.4	4.8	0.069	5.2	5.2	0.002
Antithrombotic agents	17.5	11.5	**0.169**	14.5	11.9	0.077
Diuretics	18.9	12.4	**0.182**	15.7	12.8	0.083
Agents acting on the renin angiotensin system	18.5	12.9	**0.155**	15.9	13.0	0.083
Oral corticosteroid	20.9	9.1	**0.336**	14.2	10.7	**0.107**
Inhaled corticosteroid	19.6	22.4	0.071	20.3	23.0	0.067
Disease (COPD and/or Asthma)	3.3	22.4	**0.238**	28.8	22.0	**0.157**

*IV = (PTLP): proportion of time for LABA prescription, binarized at its median, **Standardized difference in **bold** indicates confounders are not balanced between exposure as well as IV groups.

related to IV, conventional multivariable regression analysis yielded less biased estimates than IV analysis in our simulations even in the presence of a strong IV.

The bias and variation in IV estimates increased considerably when the association between IV and exposure was weak (i.e., weak IV), which is in line with previous studies.[1,2] In those cases, estimates from IV analysis were more biased than conventional regression analysis, even when measured confounders were included both in the first and second-stage IV models. However, when the IV was strong (e.g., $\beta_z = 2.5$), including measured confounders in the IV models provide essentially unbiased estimates like linear regression in the absence of unmeasured confounding despite poor performance of IV methods in finite sample size.[8]

In our empirical example, the adjusted estimate from conventional regression was not significant and close to the "no relevant differences between treatment groups" in a meta-analysis of RCTs.[45] On the other hand, the estimates from unadjusted and adjusted IV analyses were divergent which could be due to imbalance of measured confounders (age, oral corticosteroids, and disease) between IV categories, i.e., violation of the third assumption. Moreover, weak association between IV and exposure (current LABA use) was evident as reflected by wider confidence interval in the unadjusted IV estimate. Although adjustment for measured confounders in IV models improved the precision of the estimate, the IV estimate is far from the estimates in meta-analyses of RCTs on this topic.[45] This difference could be due to the imbalance in measured confounders between IV groups and it seems plausible that such imbalance could also exist in unmeasured confounders. Hence, when there is an imbalance in unmeasured confounders between IV categories, adjustment for measured confounders in the IV analysis could result in more biased estimates than conventional regression methods. Therefore, IV analysis is inappropriate in such cases. On the other hand, the non-collapsibility [46] of hazard ratio could in part explain the difference between adjusted and unadjusted (IV or conventional regression) estimates.

This study has several strengths. First, we explored the usefulness of balance measures (SDif) in several realistic settings to falsify the third assumption of IV, which is easy to apply. The SDif has several desirable properties compared to other tests of independence (e.g., t-test), including independence of sample size.[28] Second, we considered a wide range of scenarios for associations among IV, exposure, and confounders (both measured and unmeasured). A limitation of our simulation study is that we restricted the simulations to a continuous outcome with linear model. We chose this approach, because IV estimates are reported to be biased in the case of a binary outcome with non-linear model.[10,47] Future research could extend the simulations to settings with binary outcome. In addition, although the different confounders in the simulations had different associations with the exposure and/or outcome, we gave equal weights to all confounders in estimating the standardized difference. Nevertheless, the choice of the weights is not straightforward and its impact on the bias is not substantial.[28] Furthermore, we used SDif for only binary IV; however, a similar approach can be used for continuous IV, i.e., assessing balance of measured confounders between quintiles of the continuous IV.

Using Monte Carlo simulations, we demonstrated that balance measures such as the standardized difference can help researchers in assessing the third assumption of IV analysis with respect to measured confounders. Although balance on measured confounders between IV categories does not "guarantee" balance on unmeasured confounders,[14,48] despite such claims being prevalent in the medical literature in IV analysis,[15,16,19,21,23,26,49] our study indicated that balance measures seems to be useful for falsification of the IV assumption. Hence, when balance in measured confounders is achieved, investigators should rely on theoretical justifications with regard to balance of unmeasured confounders[26,50] for the validity of the IV analysis. If balance measures indicate that the confounders are imbalanced between IV categories, and thus falsify the third IV assumption, researchers should consider refraining from IV analysis even adjusting for measured confounders.

REFERENCES

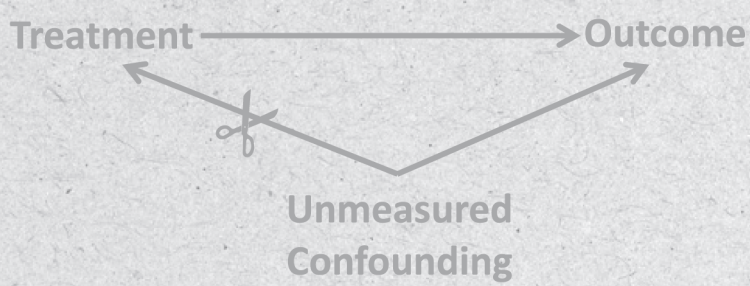
1. Martens EP, Pestman WR, De Boer A, Belitser SV, Klungel OH. Instrumental variables: Application and limitations. *Epidemiology* 2006; 17(3): 260-7.
2. Hernán MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? *Epidemiology* 2006; 17(4): 360-72.
3. Greenland S. An introduction to instrumental variables for epidemiologists. *Int J Epidemiol* 2000; 29(4): 722-9.
4. Davies NM, Smith GD, Windmeijer F, Martin RM. Issues in the reporting and conduct of instrumental variable studies: A systematic review. *Epidemiology* 2013; 24(3): 363-9.
5. Stock JH, Wright JH, Yogo M. A survey of weak instruments and weak identification in generalized method of moments. *J Bus Econ Stat* 2002; 20(4): 518-29.
6. Angrist JD, Pischke JS. Mostly harmless econometrics: An empiricist's companion. Princeton Univ Pr; 2008.
7. Staiger D, Stock JH. Instrumental variables regression with weak instruments. *Econometrica* 1997; 65(3): 557-86.
8. Bound J, Jaeger DA, Baker RM. Problems with instrumental variables estimation when the correlation between the instruments and the endogeneous explanatory variable is weak. *J Am Stat Assoc* 1995; 90(430): 443-50.
9. Yoo BK, Frick KD. The instrumental variable method to study self-selection mechanism: A case of influenza vaccination. *Value Health* 2006; 9(2): 114-22.
10. Uddin MJ, Groenwold RHH, de Boer A, Belitser S, Roes KCB, Hoes AW, Klungel OH. Performance of instrumental variable methods in cohort and nested case–control studies: a simulation study. *Pharmacoepidemiol Drug Saf* 2014; 23: 165–177.
11. Didelez V, Meng S, Sheehan NA. Assumptions of IV methods for observational epidemiology. *Stat Sci* 2010; 25(1): 22-40.
12. Chen Y, Briesacher BA. Use of instrumental variable in prescription drug research with observational data: a systematic review. *J Clin Epidemiol* 2011; 64(6): 687-700.
13. Didelez V, Sheehan N. Mendelian randomization as an instrumental variable approach to causal inference. *Stat Methods Med Res* 2007; 16(4): 309-30.
14. Glymour MM, Tchetgen EJT, Robins JM. Credible mendelian randomization studies: Approaches for evaluating the instrumental variable assumptions. *Am J Epidemiol* 2012; 175(4): 332-9.

15. 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; 54(11): 3390-8.
16. 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. *Can Med Assoc J* 2007; 176(5): 627.
17. 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; 19(4): 1064-70.
18. Wang PS, Schneeweiss S, Avorn J, Fischer MA, Mogun H, Solomon DH, Brookhart MA. Risk of death in elderly users of conventional vs. atypical antipsychotic medications. *N Engl J Med* 2005; 353(22): 2335-41.
19. Lu-Yao GL, Albertsen PC, Moore DF, Shih W, Lin Y, DiPaola RS, Yao SL. Survival following primary androgen deprivation therapy among men with localized prostate cancer. *JAMA* 2008; 300(2): 173.
20. Stuart BC, Doshi JA, Terza JV. Assessing the impact of drug use on hospital costs. *Health Serv Res* 2009; 44(1): 128-44.
21. Zhang Y. Cost-saving effects of olanzapine as long-term treatment for bipolar disorder. *J Ment Health Policy Econ* 2008; 11(3): 135-46.
22. Schneeweiss S, Seeger JD, Landon J, Walker AM. Aprotinin during coronary-artery bypass grafting and risk of death. *N Engl J Med* 2008; 358(8): 771-83.
23. Park T-, Brooks JM, Chrischilles EA, Bergus G. Estimating the effect of treatment rate changes when treatment benefits are heterogeneous: Antibiotics and otitis media. *Value in Health* 2008; 11(2): 304-14.
24. Ramirez SPB, Albert JM, Blayney MJ, et al. Rosiglitazone is associated with mortality in chronic hemodialysis patients. *Journal of the American Society of Nephrology* 2009; 20(5): 1094-101.
25. 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; 17(3): 268-75.
26. Swanson SA, Hernán MA. Commentary: How to report instrumental variable analyses (suggestions welcome). *Epidemiology* 2013; 24(3): 370-4.
27. Austin PC. The relative ability of different propensity score methods to balance measured covariates between treated and untreated subjects in observational studies. *Medical Decision Making* 2009; 29(6): 661-77.
28. Belitser SV, Martens EP, Pestman WR, Groenwold RH, de Boer A, Klungel OH. Measuring balance and model selection in propensity score methods. *Pharmacoepidemiol Drug Saf* 2011; 20(11): 1115-29.
29. Groenwold RHH, de Vries F, de Boer A, Pestman WR, Rutten FH, Hoes AW, Klungel OH. Balance measures for propensity score methods: A clinical example on beta-agonist use and the risk of myocardial infarction. *Pharmacoepidemiol Drug Saf* 2011; 20(11): 1130-7.
30. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983; 70(1): 41-55.
31. Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Stat Med* 2009; 28: 3083-107.
32. Austin PC. A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003. *Stat Med* 2008; 27(12): 2037-49.

33. Austin PC. Propensity-score matching in the cardiovascular surgery literature from 2004 to 2006: A systematic review and suggestions for improvement. *J Thorac Cardiovasc Surg* 2007; 134(5): 1128,1135.e3.
34. Austin PC. A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003. *Stat Med* 2008; 27(12): 2037-49.
35. R Development Core Team. R: A language and environment for statistical computing. *R Foundation for Statistical Computing: Vienna, Austria* 2010; ISBN 3-900051-07-0(01/19).
36. Angrist JD, Krueger AB. Instrumental variables and the search for identification: From supply and demand to natural experiments. *J Econ Perspect* 2001; 15(4): 69-85.
37. Burgess S. Statistical issues in Mendelian randomization: use of genetic instrumental variables for assessing causal associations. Diss. University of Cambridge, 2012.
38. Burgess S, Thompson SG, Andrews G, et al. Bayesian methods for meta-analysis of causal relationships estimated using genetic instrumental variables. *Stat Med* 2010; 29(12): 1298-311.
39. Brookhart MA, Rassen JA, Schneeweiss S. Instrumental variable methods in comparative safety and effectiveness research. *Pharmacoepidemiol Drug Saf* 2010; 19(6): 537-44.
40. Ali MS, Groenwold RH, Pestman WR, Belitser SV, Hoes AW, de Boer A, Klungel OH. Time-dependent propensity score and collider-stratification bias: an example of beta2-agonist use and the risk of coronary heart disease. *Eur J Epidemiol* 2013; : 1-9.
41. De Vries F, Pouwels S, Bracke M, Lammers J, Klungel O, Leufkens H, Van Staa T. Use of β2 agonists and risk of acute myocardial infarction in patients with hypertension. *Br J Clin Pharmacol* 2008; 65(4): 580-6.
42. Schmoor C, Caputo A, Schumacher M. Evidence from nonrandomized studies: A case study on the estimation of causal effects. *Am J Epidemiol* 2008; 167(9): 1120-9.
43. Bosco JLF, Silliman RA, Thwin SS, et al. A most stubborn bias: no adjustment method fully resolves confounding by indication in observational studies. *J Clin Epidemiol* 2010; 63(1): 64-74.
44. 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.
45. Ferguson GT, Funck-Brentano C, Fischer T, Darken P, Reisner C. Cardiovascular safety of salmeterol in COPD. *Chest* 2003; 123(6): 1817-24.
46. Greenland S, Robins JM, Pearl J. Confounding and collapsibility in causal inference. *Stat Sci* 1999; : 29-46.
47. Palmer TM, Thompson JR, Tobin MD, Sheehan NA, Burton PR. Adjusting for bias and unmeasured confounding in Mendelian randomization studies with binary responses. *Int J Epidemiol* 2008; 37(5): 1161-8.
48. Ward A, Johnson PJ. Addressing confounding errors when using non-experimental, observational data to make causal claims. *Synthese* 2008; 163(3): 419-32.
49. Tentori F, Albert JM, Young EW, et al. The survival advantage for haemodialysis patients taking vitamin D is questioned: Findings from the Dialysis Outcomes and Practice Patterns Study. *Nephrology Dialysis Transplantation* 2009; 24(3): 963-72.
50. Burgess S. Re: "credible mendelian randomization studies: Approaches for evaluating the instrumental variable assumptions". *Am J Epidemiol* 2012; 176(5): 456-7.

Unmeasured confounding IV PERR SCCS Observational studies
PROTECT Causal Inference Statistical methods পরিসংখ্যানগত পদ্ধতি
Statistische methoden IV PERR SCCS Pseudo Randomization
Pharmacoepidemiology IV Analyses PERR SCCS
General Practice Databases Residual Confounding
Electronic Healthcare Databases Epidemiology
Time-fixed Confounding Time-fixed Exposure
Time-varying Exposure
Time-varying Confounding
Non-randomized Studies
IV PERR SCCS
PROTECT 2SLS
IV IV

CHAPTER IV
APPLICATIONS OF INSTRUMENTAL VARIABLE ANALYSIS



CHAPTER 4.1

Evaluating Different Physician's Prescribing Preference Based Instrumental Variables in Two Primary Care Databases: a Study of Inhaled Long-acting Beta2-Agonist Use and the Risk of Myocardial Infarction

Md Jamal Uddin, Rolf HH Groenwold, Anthonius de Boer, Ana SM Afonso, Paola Primatesta, Claudia Becker, Svetlana V Belitser, Arno W Hoes, Kit CB Roes, Olaf H Klungel

Submitted, Pharmacoepidemiology and Drug Safety

ABSTRACT

BACKGROUND: Instrumental variable (IV) analysis with physician's prescribing preference (PPP) as IV is increasingly used in pharmacoepidemiology. However, it is unclear whether this IV performs consistently across databases. We aimed to evaluate the validity of different PPPs in a study of inhaled long-acting beta2-agonist (LABA) use and myocardial infarction (MI).

METHODS: Information on adults with asthma and/or COPD and at least one prescription of beta2-agonist, or muscarinic antagonist was extracted from the CPRD, UK and the Dutch Mondriaan databases. LABA exposure was considered time-fixed or time-varying. We measured PPPs using previous LABA prescriptions of physicians or proportion of LABA prescriptions per practice. Correlation (r) and standardized difference (SDif) were used to assess the validity of IVs.

RESULTS: For time-fixed LABA, the IV based on 10 previous prescriptions outperformed the other IVs regarding strength of the IV ($r \geq 0.15$) and balance of confounders between IV categories (SDif < 0.10). None of the IVs we considered appeared to be valid for time-varying LABA. In CPRD (n = 490499), which included approximately 18 times more subjects than Mondriaan (n = 27459), IVs appeared more valid. LABA was not associated with MI; hazard ratios ranged from 0.86 to 1.18 for conventional analysis, and from 0.61 to 1.24 for the IV analyses with apparent valid IVs.

CONCLUSIONS: The validity of IV PPP strongly depends on how this IV is defined and in which database it is applied. Hence, a general recommendation cannot be given, other than to generate several plausible IVs, assess their validity, and report the estimate(s) from apparently valid IVs.

INTRODUCTION

Healthcare databases often have limited information available on potential confounders, such as severity of disease,[1,2] which may result in biased estimates of adverse effects of drugs. Instrumental variable (IV) analysis has been proposed to control for unmeasured confounding, and has been applied in some pharmacoepidemiological studies.[2–10] An IV 1) is associated with the exposure, 2) affects the outcome only through the exposure, and 3) is independent of confounders.[11–13] If these key assumptions and some additional assumptions (e.g., homogeneity) are satisfied, IV analysis can consistently estimate the average causal effect (ACE) of an exposure.[13–15]

Brookhart et al.[2] have proposed the physician's prescribing preference (PPP) as an IV in 2006, and it is increasingly being used for IV analysis since then.[3,5,7,8,16,17] PPP can be defined in different ways, for example based on the proportion of prescriptions of a particular drug.[2,5,7,17,18] Previous studies have assessed different definitions of PPPs at either hospital or physician level. Examples include studies of antipsychotic medication use and risk of

death,[8,18] non-steroidal anti-inflammatory drugs use and risk of gastrointestinal tract complications and myocardial infarction.[7] Although it was acknowledged in these studies that treatment effects may differ considerably depending on the definition of PPP, it is unknown how different definitions of PPPs perform across different databases especially in the case of time-varying exposures.

Several observational studies have been conducted on the association between inhaled beta2-agonist use and myocardial infarction (MI), leading to conflicting results,[19–25] which may be due to unmeasured confounding.[25,26] We aimed to evaluate the applicability and apparent validity of different definitions of PPP at practice level in two general practice (GP) databases from the United Kingdom (UK) and the Netherlands in a study of inhaled long-acting beta-2-receptor agonist (LABA) use and the risk of MI.

METHODS

Data sources

We performed studies in the UK-based Clinical Practice Research Datalink (CPRD)[27] and the Dutch Mondriaan GP database.[28–30] Detailed information on the databases can be found in a common study protocol.[31]

Study population and cohort

We sampled 1,037,647 patients from CPRD and 116,240 patients from Mondriaan with at least one prescription of inhaled short- and/or long-acting beta-2-receptor agonist (SABA, LABA) and/or inhaled short- and/or long-acting muscarinic antagonist (SAMA, LAMA) during the study period, 1 January 2002 to 31 December 2009. Moreover, for each patient at least one year of enrolment with his/her GP was required to enter the study. We excluded patients (n = 547,148 in CPRD and n = 88,781 in Mondriaan) from the study population that were younger than 18 years (n = 166,222), had a previous MI (n = 24,678), or did not have a coded diagnosis of asthma or COPD, or used inhalation therapy for other indications (n = 445,029). Moreover, practices that contributed with less than 50 patients were also excluded, because a considerable number of prescriptions is needed to accurately estimate the IV PPP.[17] In CPRD we used 590 practices and in Mondriaan 167 practices.

The date of cohort entry was defined for each patient as the date of first prescription of at least one SABA, LABA, SAMA, or LAMA. The observation period for each patient lasted until the end of data collection (31 December 2009), the first MI, death, or unregistering, whichever occurred first.

Exposures and Outcome

We considered two types of exposure: 1) exposure that was determined at baseline and considered constant over time; and 2) exposure that was time-varying. The first exposure was defined based on the first prescription (LABA vs. SABA/SAMA/LAMA). Any changes in

medication status during follow-up (e.g. switch from SABA to LABA or stopping LABA) were ignored. Throughout the manuscript we will use the abbreviation "$LABA_{fixed}$" to indicate the exposure effect using the first prescription. As it seems unrealistic to assume that exposures and confounders remain constant throughout follow-up,[32] the second type of exposure was time-varying. The time-varying exposure was determined by calculating the length of treatment periods of LABA use. When subsequent prescription with the same drug was collected before the theoretical end date of a previous prescription, the number of overlapping days was added to the theoretical end date of the subsequent prescription. Hence, the follow-up time of exposure to LABA was divided into "current LABA use" (from the beginning of the prescription up to the calculated end date of the prescription) and "non-LABA use". Non-LABA use includes current or non-current SABA/SAMA/LAMA use or recent/past LABA use. Throughout the manuscript we will use the abbreviation "$LABA_{time}$" to indicate the exposure effect in time-varying setting. The Anatomical Therapeutic Chemical and Multilex Codes for the drug exposures of interest were described in the Appendix, Table A3.

The first fatal or non-fatal MI within the study period was the outcome of interest. See detailed International Classification of Primary Care and Read Codes for the outcome in the Appendix (Table A4) and study protocol.[31]

Potential confounding factors

The potential confounders (i.e., co-medications, co-morbidities, and lifestyle factors) are listed in the Table 1. For the exposure $LABA_{fixed}$, confounders were assessed at baseline and for the exposure $LABA_{time}$ confounders were considered time-varying and assessed at baseline and updated whenever patients' exposure status changed or every six months (if exposure status did not change). For co-morbidities, patients were classified as having the disease from the first date of diagnosis onwards.

Instrumental variables

For the exposure $LABA_{fixed}$, several IVs were generated based on the previous prescriptions (LABA vs. non-LABA) of a GP or proportion of prescriptions in a particular practice (Figure 1).[2,17] For previous prescriptions, we used a single last prescription (PPP1), last five (PPP5), and last ten prescriptions (PPP10). When the last single prescription was LABA, the IV PPP1 takes value 1 and 0 otherwise.[2] The value of the IVs PPP5 and PPP10 is the proportion of LABA prescriptions based on the last 5 or 10 prescriptions, respectively. Similarly, we also created IVs using previous 20 or 50 prescriptions. Moreover, we used the overall proportion of LABA prescriptions per practice (PLP), which was also dichotomized (PLP_{dich}) at the median.[33]

For the time-varying exposure $LABA_{time}$, the IVs were defined as the proportions of time of "current" LABA use in a particular practice. Specifically, it is the ratio of the follow-up time for "current" LABA users in a practice to the total follow-up time for all patients in the same practice. Moreover, this IV was dichotomized at the median to create a binary IV.

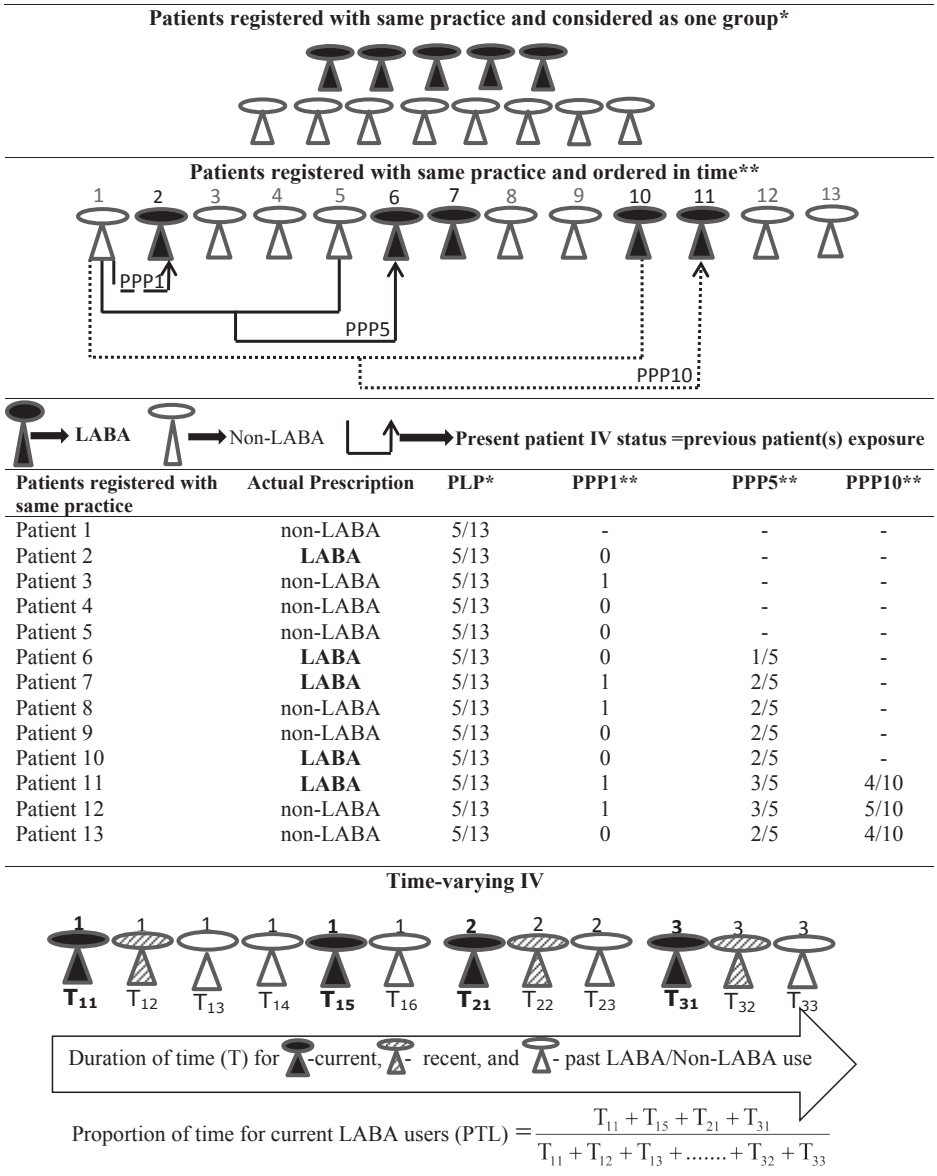
Table 1. Characteristics of patients stratified on type of beta2-agonist prescriptions (LABA vs. SABA/SAMA/LAMA) assigned at index date

	CPRD: COPD		CPRD: Asthma		CPRD: COPD & Asthma		CPRD: Combined¥±		Mondriaan: Combined¥	
	Non-LABA	LABA	Non-LABA	LABA	Non-LABA	LABA	Non-LABA	LABA	Non-LABA	LABA
Sample size	34850	6716	325520	49259	47850	16757	415470	75029	19526	7933
MI	995	167	1901	390	1594	481	4659	1080	286	161
Confounders (%)										
Age (Mean, (SD))	69 (11.4)	70 (10.7)	43 (17.2)	47 (17.2)	66 (12.5)	66 (11.8)	48 (19.2)	54 (18.3)	49 (18.0)	55 (16.8)
Sex (male)	54.1	57.8	41.1	38.4	46.9	47.4	43.1	42.6	41.6	43.5
Ischemic heart disease	14.7	14.7	3.0	4.1	13.4	12.6	5.4	7.2	3.5	4.1
Hypertension	36.0	35.6	13.3	16.9	31.9	31.5	17.7	22.4	12.4	14.5
Stroke	9.3	9.5	1.6	1.9	7.5	6.3	3.1	3.8	1.2	1.4
Thromboembolic disease	12.1	11.6	2.9	3.5	9.9	9.8	4.6	5.8	-	-
Diabetes mellitus	11.6	11.5	4.4	5.6	11.0	10.4	5.9	7.4	6.6	7.4
Hypercholesterolemia	12.5	11.4	4.6	6.0	10.3	10.0	6.1	7.5	-	-
Antithrombotic agents	31.5	34.3	6.1	8.3	23.6	22.6	10.7	14.5	10.6	12.6
Disease (COPD)	-	-	-	-	-	-	9.4	10.4	63.7	49.0
Disease (Asthma)	-	-	-	-	-	-	78.4	65.7	25.3	34.6
Beta-Blocker	14.8	12.8	3.6	2.5	6.9	3.5	5.1	3.9	9.6	11.3
Vasodilators	10.4	10.8	2.1	3.2	9.1	9.2	3.7	5.4	<0.01	0.1
Calcium channel blockers	18.9	20.1	5.6	7.7	16.9	17.4	8.3	11.3	5.2	6.1

Table 1. Characteristics of patients stratified on type of beta2-agonist prescriptions (LABA vs. SABA/SAMA/LAMA) assigned at index date (*Continued*)

	CPRD: COPD		CPRD: Asthma		CPRD: COPD & Asthma		CPRD: Combined¥±		Mondriaan: Combined¥	
	Non-LABA	LABA	Non-LABA	LABA	Non-LABA	LABA	Non-LABA	LABA	Non-LABA	LABA
Diabetes drugs*	6.0	5.9	2.7	3.5	5.7	5.4	3.4	4.3	5.5	6.1
Lipid modifying agents (e.g., statins, fibrates)	20.9	23.6	5.5	8.0	13.7	14.4	8.0	11.3	8.4	10.4
Aspirin	25.6	27.5	5.0	6.5	19.1	17.9	8.7	11.4	-	-
Diuretics**	34.3	36.8	9.5	13.1	31.3	32.5	14.5	20.3	10.7	12.9
Agents on the renin angiotensin system	19.4	20.9	5.8	7.8	15.7	16.1	8.3	11.2	10.5	13.0
Xanthines	3.2	6.5	0.9	3.3	7.5	12.7	1.9	5.9	0.7	1.6
Inhaled corticosteroid	16.7	22.5	35.2	49.6	52.1	60.6	35.5	49.3	17.4	18.8
Oral corticosteroid	14.3	24.4	8.0	22.9	25.1	40.9	10.7	27.3	7.5	12.7
Smoking	89.7	89.9	69.5	69.3	88.7	88.6	73.8	76.0	3.7	4.2
Antiallergics excl. corticosteroids	<0.01	<0.01	0.3	0.4	0.3	0.4	0.3	0.4	-	-
Smoking	89.7	89.9	69.5	69.3	88.7	88.6	73.8	76.0	3.7	4.2
Other Information										
Total follow-up (year)	143743		1688872		331113		2195280		105739	
Median follow-up (year)	3.05		4.67		5.61		4.58		3.50	
Number of practices	390		588		464		590		167	

CPRD- Clinical practice research datalink; COPD- Chronic obstructive pulmonary disease; LABA- Long acting beta2-agonist; SABA- Short acting beta2-agonist; SAMA- Short acting muscarinic antagonists; LAMA- long acting muscarinic antagonists; *Anatomical Therapeutic Chemical Code is A10. **Anatomical Therapeutic Chemical code is C03; ± sum of patients in all cohorts is not equal to the number of patients in combined cohort because less number of practices were removed for less than 50 patients.



Patients registered with same practice	Actual Prescription	PLP*	PPP1**	PPP5**	PPP10**
Patient 1	non-LABA	5/13	-	-	-
Patient 2	**LABA**	5/13	0	-	-
Patient 3	non-LABA	5/13	1	-	-
Patient 4	non-LABA	5/13	0	-	-
Patient 5	non-LABA	5/13	0	-	-
Patient 6	**LABA**	5/13	0	1/5	-
Patient 7	**LABA**	5/13	1	2/5	-
Patient 8	non-LABA	5/13	1	2/5	-
Patient 9	non-LABA	5/13	0	2/5	-
Patient 10	**LABA**	5/13	0	2/5	-
Patient 11	**LABA**	5/13	1	3/5	4/10
Patient 12	non-LABA	5/13	1	3/5	5/10
Patient 13	non-LABA	5/13	0	2/5	4/10

Figure 1. Definition of instrumental variables based on proportion of LABA prescriptions and previous prescription written by physician.

LABA: long acting beta2-agonist, PLP: proportion of LABA prescription, PPP1: physician prescribing preference with single last prescription (binary, 1 = LABA or 0 = non-LABA); PPP5: physician prescribing preference with previous five prescriptions, e.g., when patients are ordered in time, IV for patient number 6 is 1/5 (i.e., among the last 5 prescriptions of patient 6, 1 prescriptions are LABA), PPP10: physician prescribing preference with previous ten prescriptions, e.g., when patients are ordered in time, IV for patient number 11 is 4/10 (i.e., among the last 10 prescriptions of patient 11, 4 prescriptions are LABA). PPP5/PPP10 is continuous variable (0-100%) *All subjects under a practice are considered as one group; **Subjects are ordered in time by practice.

Assessment of instrumental variable assumptions

To check the first IV assumption (i.e., association between IV and exposure), we used point bi-serial correlation (r) and odds ratio (OR) when exposure was binary and IV was continuous and both exposure and IV were binary, respectively.[5,13] We assumed that a physician's preference is unlikely to affect the outcome directly (second assumption) or is associated with patient characteristics (third assumption).[17] Some authors argued that the second and third assumptions are untestable as they involve unmeasured variables.[11,14] However, we applied a falsification test for the third assumption using the standardized difference (SDif)[34] to assess balance of measured confounders between IV categories (for continuous IV, balance was assessed across the quintiles). We also used the Mahalanobis distance to measure the balance on several covariates simultaneously.[18,35] Lower values of SDif (e.g., <0.10)[36] and Mahalanobis distance indicate better balance.[34,35] If measured confounders are insufficiently balanced between IV categories, this may imply imbalance of unmeasured confounders.[34] In that case, the third assumption is violated and IV analysis appears inappropriate.[34] However, if measured confounders are balanced, we assumed that such balance could be carried-over to unmeasured confounders.[2,9]

Statistical analyses

To reduce heterogeneity between patients, we stratified CPRD patients according to their diagnosis of COPD, asthma, COPD & asthma. The limited number of cases in Mondriaan did not allow for this stratification. To allow comparison across databases, we also performed an analysis including all CPRD patients. Additionally, we analysed pooled data from the combined cohort of CPRD and Mondriaan. To reduce heterogeneity between patients in the combined cohorts, we adjusted for disease (i.e. asthma, COPD, and asthma and COPD) in the IV and conventional models.

Data were analysed using the Cox proportional hazards model and two-stage IV models (time fixed and time-varying) with and without adjustment for confounders. For the exposure $LABA_{fixed}$, the baseline confounders and for the exposure $LABA_{time}$, time-varying confounders were considered in the adjusted models. In the IV analysis, the first-stage model was a linear regression model,[37,38] in which exposure was the dependent variable and the IV was the independent variable. The second-stage model was a Cox model, in which the predicted exposure rather than actual exposure was the independent variable.[39–42] We also applied 2-stage residual inclusion method in the IV analysis. Since both the two-stage models and the 2-stage residual inclusion method provided similar results, we reported results only from the two-stage IV models. Confidence intervals (CIs) for IV estimates were estimated using bootstrapping (number of bootstrap samples 1000). All analyses were performed using the statistical software R version 2.15.2.[43]

We studied several cohorts (e.g., COPD, asthma) and assumed that patients within a cohort are homogeneous (i.e., the effect of LABA on MI is approximately the same for all subjects). Hence, this assumption allows to identify the point estimate of the ACE of the exposure

from the IV analysis.[11] Moreover, this also allows comparing the IV effect estimates with the estimates from conventional analyses.[13,44]

RESULTS

In CPRD, patients with a diagnosis of only asthma were least likely (13%) to use LABA, whereas patients with both asthma and COPD were most likely (26%) (Table 1). In total, 5,739 persons in the CPRD and 447 persons in Mondriaan experienced an MI during follow-up. When confounders were compared between exposure groups and categories of IVs, the confounders were more balanced between categories of the proposed IVs than between LABA exposure groups (Table A2a-A2c, Appendix).

The quantitative assessment of IV assumptions is summarized in Table 2, details can be found in the Appendix. The IV PLP was strongly associated with $LABA_{fixed}$ (r = 0.15, Table A1, Appendix) and confounders were balanced (SDif < 0.10, Table A2a, Appendix) in the asthma cohort of the CPRD. Similar performance was observed for the PLP_{dich} in the COPD and COPD & asthma cohorts. The IV PPP1 was weakly associated with $LABA_{fixed}$ (OR < 2 in all cohorts, Table A1, Appendix), but the confounders were balanced (SDif < 0.10, Table A2a-b, Appendix). Because of the weak association with exposure, the estimates based on PPP1 were highly unstable with very wide confidence intervals (Table 2). For example, in the asthma cohort, the difference between adjusted and unadjusted estimates was very large, unadjusted hazard ratio (HR) 35.6 [95% CI 3.86-284] and adjusted HR 7.41 [95% CI 0.77-78.2]

Table 2. Summary results from IV analysis with first prescription and time-varying setting

Database	Cohort	IVs	Assumption-1	Assumption-3	Estimates Unadj-Adj	Length CI: Adj-Model
CPRD	COPD	PLP		X	0.56-0.55	0.94
		PLP_{dich}			**0.89-0.95**	**2.39**
		PPP1	X		1.04-0.79	6.25
		PPP5	X		0.62-0.53	1.62
		PPP10			**0.72-0.53**	**1.56**
		PTL*	X	X	0.27-0.28	0.61
		PTL_{dich} *	X		0.26-0.29	1.01
	Asthma	**PLP**			**2.10-0.75**	**1.55**
		PLP_{dich}	X		1.61-0.65	1.81
		PPP1	X		35. 6-7.41	77.4
		PPP5	X		7.24-1.21	3.35
		PPP10	X		7.34-1.07	2.61
		PTL*	X		11.5-2.40	6.51
		PTL_{dich} *	X	X	10.1-2.59	12.0

Table 2. Summary results from IV analysis with first prescription and time-varying setting (*Continued*)

Database	Cohort	IVs	Assumption-1	Assumption-3	Estimates Unadj-Adj	Length CI: Adj-Model
	COPD & Asthma	PLP		X	0.64-0.52	0.57
		PLP_{dich}			**0.79-0.69**	**0.84**
		PPP1	X		1.41-0.83	3.35
		PPP5	X		1.12-0.69	1.32
		PPP10			**1.15-0.72**	**1.05**
		PTL*	X	X	0.42-0.35	0.60
		PTL_{dich}*	X		0.24-0.23	0.75
	Combined[¥]	PLP	X		1.05-0.54	0.62
		PLP_{dich}	X		1.04-0.59	0.87
		PPP1	X		17.4-0.93	3.47
		PPP5	X		8.9-0.80	1.24
		PPP10			**8.7-0.95**	**1.08**
		PTL*	X		10.5-2.18	3.57
		PTL_{dich}*	X		11.6-2.40	6.75
Mondriaan	Combined[¥]	PLP		X	1.43-0.66	1.13
		PLP_{dich}		X	1.15-0.48	1.26
		PPP1	X		4.72-1.92	10.3
		PPP5			**2.59-1.24**	**3.20**
		PPP10			**2.46-1.24**	**2.62**
		PTL*		X	1.18-0.14	0.65
		PTL_{dich}*		X	1.69-0.19	1.60

CPRD- Clinical practice research datalink; COPD- Chronic obstructive pulmonary disease; IV- instrumental variable; PLP- proportion of long acting beta2-agonist prescriptions by practice (continuous); PLP_{dich}- proportion of long acting beta2-agonist prescriptions by practice (binary); PPP1- single previous prescription by a physician; PPP5- previous five prescriptions by a physician; PPP10- previous ten prescriptions by a physician; PTL- proportions of time of "current" long acting beta2-agonist use in a particular practice (continuous); PTL_{dich}- proportions of time of "current" long acting beta2-agonist use in a particular practice (binary); "X" indicates instrumental variables assumptions are violated; Unadj- unadjusted estimates from the models that are not included confounders; Adj- adjusted estimates from the models that are included confounders; CI-confidence interval; **BOLD** lines indicate the potentials instrumental variables; [¥]COPD, Asthma, COPD and Asthma cohorts are combined together; *IVs for time-varying IV analysis.

(Table 3). On the other hand, the IV PPP10 was strongly associated with $LABA_{fixed}$ (Table A1, Appendix), except the asthma cohort, and confounders were balanced across the quintiles of the PPP10 in all cohorts (data not shown). Thus the stability and precision of the estimates based on the PPP10 were higher than the estimates based on the PPP1 (Table 2). When the number of previous prescriptions increased (e.g., previous 20 or 50 prescriptions) the IV estimates were approximately similar to those observed for the IV PPP10 (data not

Table 3. Association between beta2-agonist use (based on first prescription) and risk of myocardial infarction based on conventional and IV analysis

			Conventional[§]	Instrumental Variables				
				PLP	PLP_{dich}	PPP1	PPP5	PPP10
Database	Cohort	Estimates	HR [CI]	HR [CI]*	HR [CI] *	HR [CI] *	HR [CI] *	HR [CI] *
CPRD	COPD	Unadjusted	1.04 [0.88-1.23]	0.56 [0.24-1.20]	**0.89 [0.34-2.43]**	1.04 [0.12-7.47]	0.62 [0.18-2.04]	**0.72 [0.21-2.23]**
		Adjusted	0.97 [0.82-1.14]	0.54 [0.23-1.17]	**0.95 [0.32-2.71]**	0.79 [0.09-6.34]	0.53 [0.15-1.77]	**0.61 [0.20-1.76]**
	Asthma	Unadjusted	1.42 [1.28-1.59]	**2.10 [0.91-4.81]**	1.61 [0.54-5.27]	35.6 [3.86-284]	7.24 [2.89-19.8]	7.34 [3.27-16.3]
		Adjusted	1.06 [0.95-1.19]	**0.75 [0.29-1.84]**	0.65 [0.19-2.00]	7.41 [0.77-78.2]	1.21 [0.30-3.65]	1.07 [0.40-3.01]
	COPD & Asthma	Unadjusted	0.89 [0.80-0.99]	0.64 [0.39-1.03]	**0.79 [0.41-1.44]**	1.41 [0.35-5.63]	1.12 [0.54-2.24]	**1.15 [0.65-2.10]**
		Adjusted	0.86 [0.78-0.96]	0.52 [0.32-0.89]	**0.69 [0.38-1.22]**	0.83 [0.16-3.51]	0.69 [0.30-1.62]	**0.72 [0.36-1.41]**
	Combined[¥]	Unadjusted	1.34 [1.26-1.44]	0.96 [0.89-1.02]	1.04 [0.55-2.00]	17.4 [5.89-49.5]	8.94 [4.94-15.6]	**8.65 [5.57-13.9]**
		Adjusted	0.96 [0.89-1.02]	0.54 [0.30-0.92]	0.59 [0.31-1.18]	0.93 [0.20-3.67]	0.80 [0.37-1.61]	**0.95 [0.55-1.63]**
Mondriaan	Combined[¥]	Unadjusted	1.43 [1.18-1.73]	1.43 [0.74-2.87]	1.15 [0.49-3.23]	4.72 [0.83-22.0]	**2.59 [1.03-6.57]**	**2.46 [1.03-5.75]**
		Adjusted	1.18 [0.97-1.43]	0.66 [0.29-1.42]	0.48 [0.20-1.46]	1.92 [0.33-10.6]	**1.24 [0.40-3.60]**	**1.24 [0.47-3.09]**

CPRD- Clinical practice research datalink; COPD- Chronic obstructive pulmonary disease; PLP- proportion of long acting beta2-agonist prescriptions by practices (continuous); PLP_{dich}- proportion of long acting beta2-agonist prescriptions by practices (binary); PPP1- single previous prescription by a physician; PPP5- previous five prescriptions by a physician; PPP10- previous ten prescriptions by a physician; Unadjusted- estimates from the models that are not included confounders; Adjusted- estimates from the models that are included confounders; [¥]COPD, Asthma, COPD & Asthma cohorts are combined together; **Confounders in the adjusted models are listed in the Table 1; [§]Cox Model; CI-confidence interval; *Confidence intervals are estimated by bootstrap method. HR with **BOLD** indicates that the corresponding IV fulfilled at least two IV assumptions.

shown). When the strength of the relation between IV and LABA increased, the imbalance in confounders between IV groups increased as well. For example, PLP and PLP_{dich} in the Mondriaan were very strongly associated with LABA exposure (r = 0.31 for PLP and OR = 3.12 for PLP_{dich}, Table A1, Appendix) with $LABA_{fixed}$ and confounders were imbalanced (e.g., SDif > 0.10 for age and inhaled corticosteroids, Table A2b, Appendix). For a pooled analysis, we only considered the IV PPP10 as it appeared to perform better than other IVs. In this setting, PPP10 was strongly associated with LABA (r = 0.17) and confounders were balanced across the quintiles of the PPP10 (SDif < 0.10).

Table 3 shows the HRs of the association between $LABA_{fixed}$ and MI. Conventional analysis showed that LABA was not associated with an increased risk of MI compared to non-LABA in almost all cohorts in both databases; the adjusted HRs ranged from 0.86 to 1.18 (Table 3). IV analysis based on the valid IVs also showed no association between LABA and MI in all cohorts in both databases; the adjusted HRs ranged from 0.61 to 1.24 (Table 3). The width of the CIs for adjusted estimates ranged from 0.13 to 0.46 for conventional analysis and from 0.84 to 3.20 for IV analysis with valid IVs. The pooled analysis resulted for the conventional analysis in an unadjusted HR of 1.38 [1.30-1.47] and adjusted HR of 0.98 [0.92-1.04] and the IV analysis with PPP10 produced an unadjusted HR of 8.51 [5.95-12.1] and an adjusted HR of 1.01 [0.66-1.60].

In analyses of time-varying LABA exposure $LABA_{time}$, both IVs (proportions of time of "current" LABA use, continuous or binary) were weakly associated with LABA use (Table A1, Appendix) and in some cohorts confounders were imbalanced with these IVs (Table A2c, Appendix). Table 4 shows the HRs for the time-varying LABA exposure $LABA_{time}$. The conventional analysis showed that "current" LABA use was not associated with an increased risk of MI in all cohorts in both databases; the adjusted HRs ranged from 0.90 to 1.13 (Table 4). Estimates from the IV analysis showed very different estimates compared to the conventional analysis: adjusted HRs ranged from 0.14 to 2.59 and some confidence intervals did not include one (Table 4). However, in this setting the IV effect estimates appeared invalid as the IVs violated at least one of the assumptions (Table 2).

DISCUSSION

Both conventional and IV analyses show that LABA use was not associated with MI risk compared to non-LABA exposure, which is consistent with results of randomized controlled trials.[45–47] However, the IV effect estimates are less precise than those of conventional analysis. For the exposure $LABA_{fixed}$, the IV PPP10 performed better than the other IVs with respect to strength of the association between IV and exposure and balance of confounders between IV categories. On the other hand, neither of the IVs we considered appears to be valid in the setting of time-varying exposure. In that case, the main limitation was the weak association between IV and LABA exposure. In many situations there is a trade-off between the strength of the IV and the balance of confounders, which has been described before.[48]

Table 4. Association between beta2-agonist use (based on time-varying setting) and risk of myocardial infarction based on conventional and IV analysis

			Conventional	Instrumental Variables	
				Proportions of time of "current" LABA use (continuous)	Proportions of time of "current" LABA use (binary)
Database	Cohort	Estimates[§]	HR [CI]	HR [CI]*	HR [CI] *
CPRD	COPD	Unadjusted	1.10 [0.97-1.24]	0.27 [0.10-0.75]	0.26 [0.08-0.84]
		Adjusted	1.08 [0.95-1.22]	0.28 [0.11-0.72]	0.29 [0.08-1.09]
	Asthma	Unadjusted	1.62 [1.47-1.78]	11.5 [3.34-30.5]	10.1 [2.29-47.3]
		Adjusted	1.11 [1.00-1.23]	2.40 [0.68-7.19]	2.59 [0.61-12.7]
	COPD & Asthma	Unadjusted	0.94 [0.86-1.02]	0.42 [0.19-0.90]	0.24 [0.08-0.70]
		Adjusted	0.94 [0.86-1.03]	0.35 [0.16-0.76]	0.23 [0.07-0.82]
	Combined[¥]	Unadjusted	1.82 [1.72-1.93]	10.5 [4.76-20.2]	11.6 [4.98-34.8]
		Adjusted	1.13 [1.07-1.20]	2.18 [1.01-4.58]	2.40 [1.04-7.79]
Mondriaan	Combined[¥]	Unadjusted	1.39 [1.15-1.69]	1.18 [0.29-4.71]	1.69 [0.34-9.40]
		Adjusted	0.90 [0.73-1.10]	0.14 [0.03-0.68]	0.19 [0.02-1.62]

CPRD- Clinical practice research datalink; COPD- Chronic obstructive pulmonary disease; [¥]COPD, Asthma, COPD and Asthma cohorts are combined together; LABA- long acting beta2-agonist; *CI- Confidence Intervals estimated by bootstrap method; **Confounders in the adjusted models are listed in the Table 1; [§]Time dependent Cox Model

We found that the conventional analysis provides consistent results for both time-fixed and time-varying exposure across the databases; however, IV analysis provides different results, which is due to the violation of crucial IV assumptions.

On the basis of our assessment criteria, we did not identify any IV that is consistently valid across the cohorts and databases, even though the IV PPP10 was strongly associated with LABA exposure ($LABA_{fixed}$) in almost all cohorts (except for the asthma cohort) and observed confounders were balanced across the quintiles of the PPP10 in all cohorts. This indicates that an IV based on multiple previous prescriptions seems to perform better than an IV based on a single prescription, which is in line with previous studies.[17,18] The optimal number of prescriptions included in the IV may differ between studies. Obviously, choices on the number of prescriptions used for the IV should depend on the IV assumptions, not on the IV effect estimate.

In some situations (e.g., analysis of the PPP10 in combined cohorts), we observed that the unadjusted and adjusted IV estimates were very different event though the IV assumptions appeared to hold. This could be due to unmeasured confounding which is related to measured confounders or measured confounders which are not really independent of the IV.

All proposed IVs for $LABA_{time}$ violated at least one of the assumptions. In CPRD, both IVs considered were weakly associated with the exposure $LABA_{time}$, which means that apparently there is a little variation in prescribing patterns between practices. In Mondriaan, both IVs were strongly associated with $LABA_{time}$, but measured confounders were imbalanced between IVs categories. Hence, it appears difficult to define a valid IV in a situation when exposure and confounders are time-varying, which is in line with the literature.[11,14,49]

We observed that when an IV is weakly associated with the exposure (e.g., PPP1 in the asthma cohort), the exposure effect was highly variable with wide confidence intervals and estimates were uninformative. Moreover, such a weak IV may amplify bias due to small violations of second and third assumptions. Consequently, IV estimates will be more biased than the conventional estimates.[5,11,14,50] To identify a weak IV, we used an arbitrary cut-off ($r < 0.15$ and $OR < 2$), which may not be appropriate for other studies.[13]

We assumed that if observed confounders were sufficiently balanced between IV categories, indicated by SDif values that are close to zero, unmeasured confounders were also balanced (third IV assumptions).[2,7,9] Additionally, we assumed physicians to act completely the same (equal standard of care) except for the preference to prescribe LABA, which implies similar concomitant medication prescribing behaviour. However, this assumption may not hold, given that LABA preference may differ between physicians.

Our study has several strengths. We applied IV analysis with a very large cohorts (especially for CPRD) that may overcome the finite-sample bias and to our knowledge this is the first IV analysis for time fixed and time-varying LABA exposure and MI risk using multiple IVs in different databases (based on the same study protocol). There are some limitations to our study. First, although we applied a robust balance measure (SDif) to falsify the third assumption,[36,51] this could fail to falsify an IV even when the third assumption is violated.[14] Second, we found that the PLP appears a valid IV in some situations. However, if physician preference is thought to change during the follow-up, PLP may not be a valid IV.[2] Third, because all IVs in the time-varying setting ($LABA_{time}$) yielded estimates that were not valid, we cannot conclude from IV analysis that there is no difference between the association of LABA and non-LABA exposure on the risk of MI. Finally, we have only information on prescriptions; however, we do not know whether patients actually used their medicines.

Several recommendations can be made based on this study. First, to evade the uncertainty of the IV effect estimates based on a single IV, it is useful to start IV analysis with different plausible IVs. Second, if there is any evidence from a verification (e.g., correlation or odds ratio for first assumption) or falsification test (e.g., SDif or Mahalanobis distance for second

assumption) that IV assumptions are violated, the IV should be treated as invalid and one should not proceed with IV analysis. Alternatively, if an IV satisfies all assumptions, the modelling assumptions should be justified and any inconsistency of the modelling assumptions reported, which helps to understand the validity of the IV estimates.

In conclusion, we explored the validity and applicability of several physicians' prescribing preference based IVs at general practice level in a pharmacoepidemiological study. The validity of IV analysis strongly depends on how this IV is defined and in which database it is applied. It remains challenging to obtain a valid IV in pharmacoepidemiological studies especially in a setting of time-varying exposure and confounders. We recommend researchers to construct several plausible IVs, to assess their validity by using available methods and clinical knowledge, and to report only estimate(s) based on those IVs that appear valid.

REFERENCES

1. Groenwold RHH, Hoes A, Nichol KL, Hak E. Quantifying the potential role of unmeasured confounders: The example of influenza vaccination. *Int J Epidemiol* 2008; 37(6): 1422-9.
2. 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; 17(3): 268-75.
3. Davies NM, Smith GD, Windmeijer F, Martin RM. Issues in the reporting and conduct of instrumental variable studies: A systematic review. *Epidemiology* 2013; 24(3): 363-9.
4. Rassen JA, Schneeweiss S, Glynn RJ, Mittleman MA, Brookhart MA. Instrumental variable analysis for estimation of treatment effects with dichotomous outcomes. *Am J Epidemiol* 2009; 169(3): 273-84.
5. Ionescu-Ittu R, Abrahamowicz M, Pilote L. Treatment effect estimates varied depending on the definition of the provider prescribing preference-based instrumental variables. *J Clin Epidemiol* 2012; 65(2): 155-62.
6. 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; 45(10): S116-22.
7. Davies NM, Smith GD, Windmeijer F, Martin RM. COX-2 selective nonsteroidal anti-inflammatory drugs and risk of gastrointestinal tract complications and myocardial infarction: An instrumental variable analysis. *Epidemiology* 2013; 24(3): 352-62.
8. Pratt N, Roughead EE, Ryan P, Salter A. Antipsychotics and the risk of death in the elderly: An instrumental variable analysis using two preference based instruments. *Pharmacoepidemiol Drug Saf* 2010; 19(7): 699-707.
9. 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; 54(11): 3390-8.
10. Huybrechts KF, Gerhard T, Franklin JM, Levin R, Crystal S, Schneeweiss S. Instrumental variable applications using nursing home prescribing preferences in comparative effectiveness research. *Pharmacoepidemiol Drug Saf* 2014; 23(8): 830-8. doi:10.1002/pds.3611.
11. Hernán MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? *Epidemiology* 2006; 17(4): 360-72.
12. Groenwold RHH, Hak E, Klungel OH, Hoes AW. Instrumental Variables in Influenza Vaccination Studies: Mission Impossible?! *Value Health* 2010; 13(1): 132-7.

13. Uddin MJ, Groenwold RHH, de Boer A, Belitser S, Roes KCB, Hoes AW, Klungel OH. Performance of instrumental variable methods in cohort and nested case–control studies: a simulation study. *Pharmacoepidemiol Drug Saf* 2014; 23: 165–177.
14. Swanson SA, Hernán MA. Commentary: How to report instrumental variable analyses (suggestions welcome). *Epidemiology* 2013; 24(3): 370-4.
15. Palmer TM, Sterne JAC, Harbord RM, et al. Instrumental variable estimation of causal risk ratios and causal odds ratios in mendelian randomization analyses. *Am J Epidemiol* 2011; 173(12): 1392-403.
16. Hennessy S, Leonard CE, Palumbo CM, Shi X, Ten Have TR. Instantaneous preference was a stronger instrumental variable than 3- and 6-month prescribing preference for NSAIDs. *J Clin Epidemiol* 2008; 61(12): 1285-8.
17. Davies NM, Gunnell D, Thomas KH, Metcalfe C, Windmeijer F, Martin RM. Physicians' prescribing preferences were a potential instrument for patients' actual prescriptions of antidepressants. *J Clin Epidemiol* 2013; 66(12): 1386-96.
18. Rassen JA, Brookhart MA, Glynn RJ, Mittleman MA, Schneeweiss S. Instrumental variables II: instrumental variable application-in 25 variations, the physician prescribing preference generally was strong and reduced covariate imbalance. *J Clin Epidemiol* 2009; 62(12): 1233-41.
19. Au DH, Curtis JR, Every NR, McDonell MB, Fihn SD. Association between inhaled ß-agonists and the risk of unstable angina and myocardial infarction. *Chest* 2002; 121(3 SUPPL.): 846-51.
20. Au DH, Lemaitre RN, Curtis JR, Smith NL, Psaty BM. The risk of myocardial infarction associated with inhaled ß- adrenoceptor agonists. *American Journal of Respiratory and Critical Care Medicine* 2000; 161(3 I): 827-30.
21. Sears MR. Adverse effects of ß-agonists. *J Allergy Clin Immunol* 2002; 110(6 SUPPL.): S322-8.
22. Cates CJ, Cates MJ. Regular treatment with formoterol for chronic asthma: serious adverse events. *Cochrane Database Syst Rev* 2012; 4:CD006923.
23. Suissa S, Assimes T, Brassard P, Ernst P. Inhaled corticosteroid use in asthma and the prevention of myocardial infarction. *Am J Med* 2003; 115(5): 377-81.
24. De Vries F, Pouwels S, Bracke M, Lammers J, Klungel O, Leufkens H, Van Staa T. Use of β2 agonists and risk of acute myocardial infarction in patients with hypertension. *Br J Clin Pharmacol* 2008; 65(4): 580-6.
25. Zhang B, De Vries F, Setakis E, Van Staa T-. The pattern of risk of myocardial infarction in patients taking asthma medication: A study with the General Practice Research Database. *J Hypertens* 2009; 27(7): 1485-92.
26. De Vries F, Setakis E, Zhang B, Van Staa TP. Long-acting ß2-agonists in adult asthma and the pattern of risk of death and severe asthma outcomes: A study using the GPRD. *European Respiratory Journal* 2010; 36(3): 494-502.
27. The Medicines and Healthcare products Regulatory Agency. The Clinical Practice Research Datalink London [cited 2014 January, 8]. Available from: http://www.cprd.com/intro.asp.
28. Julius Clinical Primary Care Research Network. [cited 2014 January, 8]. Available from: http://www.juliusclinical.com/global-presence - cat10.
29. Top Institute Pharma TN. [cited 2014 Januar, 8]. Available from: http://www.tipharma.com/projects/efficiency-analysis-drug-discovery-process/the-mondriaan-project.html.
30. Stirbu-Wagner I, Dorsman S, Visscher S, et al. Landelijk Informatienetwerk Huisartsenzorg. Feiten en cijfers over huisartsenzorg in Nederland [Netherlands Information Network of General Practice. Facts and figures about general practice care the Netherlands]. *Utrecht/Nijmegen: NIVEL/IQ* 2010.
31. ENCePP. European Network of Centres for Pharmacoepidemiology and Pharmacovigilance 2014 [2014/07/16]. Available from: http://www.encepp.eu/encepp/openAttachment/fullProtocolLatest/3690.

32. Arbogast P, VanderWeele T. Considerations for statistical analysis. In: Velentgas P and Dreyer NA, eds. Developing a Protocol for Observational Comparative Effectiveness Research: A User's GuideMD: Agency for Healthcare Research and Quality; January 2013. p. 135-144. AHRQ Publication No. 12-EHC099. Available at www.effectivehealthcare.ahrq.gov/Methods-OCER.cfm.
33. Baiocchi M, Cheng J, Small DS. Instrumental variable methods for causal inference. *Stat Med* 2014; 15(33(13)): 2297-340. doi:10.1002/sim.6128.
34. Ali M, Uddin MJ, Groenwold RHH, et al. Quantitative Falsification of Instrumental Variables Assumption Using Balance Measures. *Epidemiology* 2014; 25(5): 770-2.
35. Franklin JM, Rassen JA, Ackermann D, Bartels DB, Schneeweiss S. Metrics for covariate balance in cohort studies of causal effects. *Stat Med* 2013; 33(10): 1685-99.
36. Austin PC. A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003. *Stat Med* 2008; 27(12): 2037-49.
37. Angrist JD, Krueger AB. Instrumental variables and the search for identification: From supply and demand to natural experiments. *J Econ Perspect* 2001; 15(4): 69-85.
38. Angrist JD. Estimation of limited dependent variable models with dummy endogenous regressors. *J Bus Econ Stat* 2001; 19(1): 2-28.
39. Schmoor C, Caputo A, Schumacher M. Evidence from nonrandomized studies: A case study on the estimation of causal effects. *Am J Epidemiol* 2008; 167(9): 1120-9.
40. Bosco JLF, Silliman RA, Thwin SS, et al. A most stubborn bias: no adjustment method fully resolves confounding by indication in observational studies. *J Clin Epidemiol* 2010; 63(1): 64-74.
41. 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.
42. Parmar AD, Sheffield KM, Han Y, et al. Evaluating comparative effectiveness with observational data: Endoscopic ultrasound and survival in pancreatic cancer. *Cancer* 2013; 119(21): 3861-9.
43. R Development Core Team. R: A language and environment for statistical computing. *R Foundation for Statistical Computing: Vienna, Austria* 2010; ISBN 3-900051-07-0(01/19).
44. Fang G, Brooks JM, Chrischilles EA. Apples and oranges? Interpretations of risk adjustment and instrumental variable estimates of intended treatment effects using observational data. *Am J Epidemiol* 2011; 175(1): 60-5. doi:10.1093/aje/kwr283.
45. Ferguson GT, Funck-Brentano C, Fischer T, Darken P, Reisner C. Cardiovascular safety of salmeterol in COPD. *Chest* 2003; 123(6): 1817-24.
46. Iftikhar IH, Imtiaz M, Brett AS, Amrol DJ. Cardiovascular safety of long acting beta agonist-inhaled corticosteroid combination products in adult patients with asthma: A systematic review. *Lung* 2014; 192(1): 47-54.
47. Sears MR, Radner F. Safety of formoterol in asthma clinical trials: An update. *European Respiratory Journal* 2014; 43(1): 103-14.
48. Martens EP, Pestman WR, De Boer A, Belitser SV, Klungel OH. Instrumental variables: Application and limitations. *Epidemiology* 2006; 17(3): 260-7.
49. Brookhart MA, Rassen JA, Schneeweiss S. Instrumental variable methods in comparative safety and effectiveness research. *Pharmacoepidemiol Drug Saf* 2010; 19(6): 537-44.
50. Bound J, Jaeger DA, Baker RM. Problems with instrumental variables estimation when the correlation between the instruments and the endogeneous explanatory variable is weak. *J Am Stat Assoc* 1995; 90(430): 443-50.
51. Belitser SV, Martens EP, Pestman WR, Groenwold RH, de Boer A, Klungel OH. Measuring balance and model selection in propensity score methods. *Pharmacoepidemiol Drug Saf* 2011; 20(11): 1115-29.

APPENDIX

Table A1. IV assumption-1: Association between IV and long acting beta2-agonist (LABA)

Instrumental variables	Measures of strength	CPRD				Mondriaan
		COPD	Asthma	COPD & Asthma	Combined¥	Combined¥
Cohort: first prescription						
Proportion of LABA prescriptions (continuous)	Correlation	0.21	0.15	0.20	**0.14**	0.31
Proportion of LABA prescriptions (binary)	Odds Ratio	2.39	**1.90**	2.08	**1.83**	3.12
Previous 1 prescription (PPP1)	Odds Ratio	**1.72**	**1.54**	**1.43**	**1.55**	**1.78**
Previous 5 prescriptions (PPP5)	Correlation	**0.14**	**0.11**	**0.14**	**0.12**	0.22
Previous 10 prescriptions (PPP10)	Correlation	0.17	**0.14**	0.18	0.15	0.26
Cohort: time-varying settings						
Proportions of time of "current" LABA use (continuous)*	Correlation	**0.14**	**0.10**	**0.11**	**0.09**	0.16
Proportions of time of "current" LABA use (binary)*	Odds Ratio	**1.67**	**1.60**	**1. 42**	**1.47**	2.79

CPRD- Clinical practice research datalink; COPD- Chronic obstructive pulmonary disease; **Bold** numbers indicate the weak instrumental variables (weakly associated with the LABA); ¥COPD, Asthma, COPD and Asthma cohorts are combined together; *IVs for time-varying IV analysis.

Table A2a. Values of standardized difference for checking IV assumption-3: Independence between IV and Confounders

	CPRD: COPD[1]				CPRD: Asthma[1]				CPRD: COPD & Asthma[1]			
	Exposure	Instrumental variables			Exposure	Instrumental variables			Exposure	Instrumental variables		
Confounders	LABA vs. NO-LABA	PPP1*	PLP[2]	PLP_{dich}	LABA vs. NO-LABA	PPP1	PLP[2]	PLP_{dich}	LABA vs. NO-LABA	PPP1	PLP[2]	PLP_{dich}
Age (Mean)	0.079	0.032	**0.133**	0.038	**0.268**	0.005	0.036	0.050	0.008	0.001	**0.147**	0.071
Sex	0.073	0.002	0.051	0.028	0.055	0.003	0.003	0.001	0.009	0.011	0.040	0.004
Ischemic heart disease	0.000	0.007	0.011	0.041	0.057	0.005	0.009	0.000	0.024	0.006	0.064	0.009
Hypertension	0.007	0.021	0.043	0.012	**0.103**	0.010	0.031	0.027	0.008	0.010	0.006	0.024
Stroke	0.006	0.022	0.028	0.023	0.022	0.000	0.004	0.005	0.046	0.004	0.002	0.004
Thromboembolic disease	0.015	0.014	0.031	0.059	0.036	0.005	0.002	0.01	0.003	0.015	0.002	0.014
Diabetes mellitus	0.004	0.028	0.005	0.029	0.057	0.004	0.014	0.003	0.019	0.019	0.043	0.006
Hypercholesterolemia	0.033	0.010	0.069	0.053	0.06	0.002	0.065	0.001	0.008	0.004	0.042	0.023
Antithrombotic agents	0.060	0.012	0.044	0.015	0.084	0.011	0.036	0.029	0.025	0.014	0.017	0.024
Beta-Blocker	0.059	0.006	0.056	0.038	0.062	0.009	0.007	0.004	0.155	0.023	0.016	0.024
Vasodilators	0.014	0.011	0.064	0.028	0.068	0.004	0.02	0.017	0.003	0.005	0.060	0.002
Calcium channel blockers	0.031	0.015	0.003	0.010	0.085	0.003	0.04	0.024	0.014	0.013	0.010	0.012
Diabetes drugs	0.004	0.015	0.004	0.007	0.046	0.008	0.016	0.003	0.012	0.018	0.010	0.005
Lipid modifying agents	0.065	0.009	0.039	0.005	0.099	0.015	0.068	0.046	0.02	0.013	0.012	0.037
Aspirin	0.044	0.020	0.039	0.024	0.064	0.007	0.036	0.025	0.029	0.017	0.038	0.016
Diuretics	0.054	0.026	0.021	0.019	**0.115**	0.008	0.028	0.033	0.026	0.000	0.007	0.016
Renin-angiotensin-system (RAS)-acting agents	0.036	0.003	0.025	0.025	0.079	0.012	0.013	0.021	0.013	0.006	0.000	0.011
Xanthines	**0.157**	0.004	0.044	0.020	**0.171**	0.003	0.052	0.018	**0.174**	0.000	0.009	0.046
Inhaled corticosteroid	**0.146**	0.034	0.085	0.017	**0.295**	0.043	0.057	0.006	**0.172**	0.016	0.059	0.036

Table A2a. Values of standardized difference for checking IV assumption-3: Independence between IV and Confounders (*Continued*)

	CPRD: COPD[1]				CPRD: Asthma[1]				CPRD: COPD & Asthma[1]			
	Exposure	Instrumental variables			Exposure	Instrumental variables			Exposure	Instrumental variables		
Confounders	LABA vs. NO-LABA	PPP1*	PLP[2]	PLP_{dich}	LABA vs. NO-LABA	PPP1	PLP[2]	PLP_{dich}	LABA vs. NO-LABA	PPP1	PLP[2]	PLP_{dich}
Oral corticosteroid	**0.259**	0.002	0.081	0.046	**0.419**	0.000	0.023	0.023	**0.340**	0.012	0.010	0.020
Smoking	0.006	0.015	**0.130**	0.082	0.005	0.037	0.062	0.087	0.003	0.000	0.046	0.049
Antiallergics	0.004	0.004	0.018	0.012	0.018	0.000	0.016	0.006	0.018	0.003	0.028	0.002
Mahalanobis Balance	0.132	0.007	0.070	0.025	0.364	0.004	0.018	0.016	0.165	0.003	0.047	0.018

[1]Data based on the first prescription; [2]Two extreme quintiles; CPRD- Clinical practice research datalink; COPD- Chronic obstructive pulmonary disease; PLP- proportion of long acting beta2-agonist prescriptions by practice (continuous); PLP_{dich}- proportion of long acting beta2-agonist prescriptions by practice (binary); PPP1- single previous prescription by a physician; *Confounders are also balanced with PPP5- previous five prescriptions by a physician and PPP10- previous ten prescriptions by a physician; **Bold** numbers indicate the violation of IV assumption-3.

Table A2b. Values of standardized difference for checking IV assumption-3: Independence between IV and Confounders

	CPRD: Combined[1¥]				Mondriaan: Combined[1¥]			
	Exposure	Instrumental variables			Exposure	Instrumental variables		
Confounders	LABA vs. NO-LABA	PPP1	PLP[2]	PLP_{dich}	LABA vs. NO-LABA	PPP1	PLP[2]	PLP_{dich}
Age (Mean)	**0.328**	0.001	0.061	0.046	**0.326**	0.069	**0.219**	**0.193**
Sex	0.010	0.001	0.014	0.000	0.038	0.028	0.07	0.042
Ischemic heart disease	0.075	0.003	0.021	0.003	0.033	0.008	0.024	0.027
Hypertension	**0.118**	0.006	0.040	0.027	0.062	0.012	0.032	0.024
Stroke	0.038	0.003	0.008	0.008	0.016	0.005	0.001	0.017
Thromboembolic disease	0.056	0.000	0.013	0.008	-	-	-	-
Diabetes mellitus	0.062	0.001	0.009	0.005	0.033	0.002	0.011	0.007
Hypercholesterolemia	0.058	0.013	0.074	0.007	-	-	-	-
Antithrombotic agents	**0.114**	0.014	0.049	0.024	0.062	0.012	0.077	0.032
Beta-Blocker	0.060	0.019	0.013	0.002	0.058	0.023	0.025	0.040
Vasodilators	0.080	0.005	0.021	0.013	0.002	0.010	0.019	0.016
Calcium channel blockers	**0.104**	0.003	0.047	0.02	0.035	0.007	0.015	0.023
Diabetes drugs	0.045	0.005	0.018	0.007	0.026	0.001	0.008	0.021
Lipid modifying agents	**0.110**	0.036	0.074	0.044	0.067	0.015	0.069	0.046
Aspirin	0.092	0.014	0.046	0.019	-	-	-	-
Diuretics	**0.153**	0.005	0.039	0.028	0.070	0.006	0.006	0.029
Renin-angiotensin-system (RAS)-acting agents	0.098	0.007	0.023	0.025	0.079	0.024	0.077	0.056
Xanthines	**0.207**	0.002	0.052	0.010	0.080	0.007	0.010	0.020
Inhaled corticosteroid	**0.281**	0.062	0.047	0.014	0.036	0.056	**0.137**	0.061

Table A2b. Values of standardized difference for checking IV assumption-3: Independence between IV and Confounders (*Continued*)

	CPRD: Combined[1¥]				Mondriaan: Combined[1¥]			
	Exposure	Instrumental variables			Exposure	Instrumental variables		
Confounders	LABA vs. NO-LABA	PPP1	PLP[2]	PLP_{dich}	LABA vs. NO-LABA	PPP1	PLP[2]	PLP_{dich}
Oral corticosteroid	**0.433**	0.005	0.002	0.017	**0.174**	0.007	0.023	0.071
Smoking	0.052	0.029	0.058	0.096	0.031	0.012	0.008	0.045
Antiallergics	0.016	0.000	0.015	0.009	-	-	-	-
Disease (for combined cohort)	**0.309**	0.014	0.060	0.024	**0.300**	0.082	**0.197**	**0.244**
Mahalanobis Balance	0.363	0.006	0.020	0.018	0.148	0.013	0.010	0.080

[1]Data based on the first prescription; [2]Two extreme quintiles; CPRD- Clinical practice research datalink; COPD- Chronic obstructive pulmonary disease; PLP- proportion of long acting beta2-agonist prescriptions by practice (continuous); PLP_{dich}- proportion of long acting beta2-agonist prescriptions by practice (binary); PPP1- single previous prescription by a physician; *Confounders are also balanced with PPP5- previous five prescriptions by a physician and PPP10- previous ten prescriptions by a physician; **Bold** numbers indicate the violation of IV assumption-3. ¥COPD, Asthma, COPD and Asthma cohorts are combined together

Table A2c. Values of standardized difference for checking IV assumption-3: Independence between IV and Confounders

	CPRD: COPD[1]		CPRD: Asthma[1]		CPRD: COPD & Asthma[1]		CPRD: Combined[1]		Mondriaan: Combined[1]	
	IV		IV		IV		IV		IV	
Confounders	PTL^2	PTL_{dich}	PTL^2	PTL_{dich}	PTL^2	PTL_{dich}	PTL^2	PTL_{dich}	PTL^2	PTL_{dich}
Age (Mean)	**0.131**	0.004	0.022	0.056	**0.137**	0.051	0.024	0.071	**0.132**	**0.159**
Sex	0.093	0.030	0.011	0.004	0.061	0.013	0.012	0.003	0.059	0.017
Ischemic heart disease	0.038	0.002	0.005	0.015	0.07	0.025	0.024	0.039	0.008	0.002
Hypertension	0.032	0.050	0.038	0.029	0.041	0.015	0.003	0.027	0.002	0.045
Stroke	0.012	0.028	0.001	0.000	0.023	0.003	0.008	0.005	0.017	0.001
Thromboembolic disease	0.023	0.053	0.006	0.004	0.052	0.004	0.006	0.005	-	-
Diabetes mellitus	0.037	0.076	0.006	0.005	0.007	0.015	0.014	0.009	0.044	0.039
Hypercholesterolemia	0.037	0.054	0.010	0.009	0.034	0.042	0.036	0.007	-	-
Antithrombotic agents (e.g., ASA, clopidogrel)	0.001	0.015	0.009	0.034	0.002	0.000	0.023	0.052	0.058	0.077
Beta-Blocker	0.038	0.047	0.003	0.006	0.000	0.03	0.005	0.01	0.093	0.060
Vasodilators	0.071	0.000	0.008	0.029	0.034	0.003	0.019	0.042	0.018	0.004
Calcium channel blockers	0.008	0.029	0.004	0.023	0.002	0.007	0.008	0.033	0.022	0.061
Diabetes drugs	0.003	0.012	0.005	0.002	0.013	0.014	0.005	0.009	0.077	0.061
Lipid modifying agents	0.004	0.006	0.007	0.031	0.032	0.009	0.014	0.047	0.036	0.082
Aspirin	0.003	0.016	0.013	0.030	0.017	0.003	0.025	0.049	-	-
Diuretics	0.011	0.010	0.018	0.044	0.003	0.000	0.007	0.054	0.057	0.083
Renin-angiotensin-system (RAS)-acting agents	0.046	0.004	0.033	0.020	0.008	0.005	0.012	0.037	0.077	0.083
Xanthines	0.072	0.016	0.000	0.029	0.014	0.029	0.011	0.027	0.057	0.029
Inhaled corticosteroid	**0.120**	0.027	0.019	0.012	0.082	0.056	0.006	0.028	**0.135**	0.067

Table A2c. Values of standardized difference for checking IV assumption-3: Independence between IV and Confounders (*Continued*)

	CPRD: COPD[1]		CPRD: Asthma[1]		CPRD: COPD & Asthma[1]		CPRD: Combined[1]		Mondriaan: Combined[1]	
	IV		IV		IV		IV		IV	
Confounders	PTL^2	PTL_{dich}	PTL^2	PTL_{dich}	PTL^2	PTL_{dich}	PTL^2	PTL_{dich}	PTL^2	PTL_{dich}
Oral corticosteroid	0.086	0.017	0.024	0.034	0.015	0.009	0.002	0.029	0.063	**0.107**
Smoking	0.094	0.098	0.083	**0.102**	**0.142**	0.008	0.000	0.060	**0.131**	0.082
Antiallergics	0.018	0.011	0.009	0.004	0.028	0.002	0.005	0.000	-	-
Disease (for combined) cohort)	-	-	-	-	-	-	0.055	0.071	**0.268**	**0.205**
Mahalanobis Balance	0.091	0.026	0.021	0.020	0.069	0.012	0.019	0.017	0.022	0.061

[1]Data based on the time-varying settings; [2]Two extreme quintiles; IV-Instrumental variables; PTL- Proportions of time of "current" long acting beta2-agonist use (continuous); PTL_{dich} - Proportions of time of "current" long acting beta2-agonist use (binary); CPRD- Clinical practice research datalink; COPD- Chronic obstructive pulmonary disease; **Bold** numbers indicate the violation of IV assumption-3. [¥]COPD, Asthma, COPD and Asthma cohorts are combined together

Table A3a. Anatomical therapeutic chemical (ATC) codes for exposure

	Drug class	ATC code	Compound name
Exposure of interest	Inhaled LABA	R03AC12	Salmeterol
		R03AC13	Formoterol
	Inhaled LABA combinations	R03AK06	Salmeterol and other drugs for obstructive airway diseases
		R03AK07	Formoterol and other drugs for obstructive airway diseases
		R03AK27	Formoterol and Beclometasone
		R03AK28	Formoterol and Budesonide
Control drugs	Inhaled SABA	R03AC02	Salbutamol
		R03AC03	Terbutaline
		R03AC04	Fenoterol
		R03AC05	Rimiterol
		R03AC06	Hexoprenaline
		R03AC07	Isoetarine
		R03AC08	Pirbuterol
		R03AC09	Tretoquinol
		R03AC10	Carbuterol
		R03AC11	Tulobuterol
		R03AC14	Clenbuterol
		R03AC15	Reproterol
		R03AC16	Procaterol
		R03AC17	Bitolterol
	Inhaled SABA combinations	R03AK03	Fenoterol and other drugs for obstructive airway diseases
		R03AK04	Salbutamol and other drugs for obstructive airway diseases
		R03AK05	Reproterol and other drugs for obstructive airway diseases
	Inhaled SAMA	R03BB01	Ipratropium bromide
	Inhaled LAMA	R03BB02	Oxitropium bromide
		R03BB04	Tiotropium bromide

LABA: Long acting beta2-agonist; SABA: Short acting beta2-agonist; SAMA: Short-acting muscarinic antagonist; LAMA: Long-acting muscarinic antagonist.

Table A3b. Multilex codes for the drug exposures

LABA:
m9056001 FORADIL caps (for inh) 12mcg [NOV/CIBA] formoterol
m9057001 formoterol fumarate caps (for inh) 12mcg formoterol
m11509001 formoterol fumarate breath actuated inh. 6 mcg formoterol
m11509002 formoterol fumarate breath actuated inh. 12mcg formoterol
m11511001 OXIS 6 TURBOHALER 6 mcg [ASTRAZENEC] formoterol
m11512001 OXIS 12 TURBOHALER 12mcg [ASTRAZENEC] formoterol
m13469001 formoterol fumarate cfc free inh. 12mcg formoterol
m13470001 ATIMOS MODULITE cfc free inh. 12mcg [CHIESI] formoterol
m15010001 EASYHALER FORMOTEROL inh. 12mcg [ORION] formoterol
m6817001 salmeterol aerosol inh. 25mcg salmeterol
m6817002 salmeterol disc 50mcg salmeterol
m6817003 salmeterol dry powder inh. 50mcg salmeterol
m6818001 SEREVENT aerosol inh. 25mcg/actuation [GLAXO] salmeterol
m6818002 SEREVENT DISKHALER 50mcg [GLAXO] salmeterol
m6818003 SEREVENT ACCUHALER 50mcg/actuation [GLAXO] salmeterol
m13678001 salmeterol cfc free inh. 25mcg/actuation salmeterol
m13679001 SEREVENT EVOHALER cfc free inh. 25mcg [GLAXO] salmeterol
m15084001 salmeterol inh powder blisters with device 50mcg salmeterol
m15087001 salmeterol inh powder blisters (refill) 50mcg salmeterol
m15088001 SEREVENT inh powder 50mcg [GLAXO] salmeterol
m15091001 SEREVENT (REFILL) inh powder 50mcg [GLAXO] salmeterol
m15760001 beclometasone with formoterol inh. 100mcg + 6mcg beclometasone /formoterol
m15761001 FOSTAIR inh. 100mcg + 6mcg [CHIESI] beclometasone /formoterol
m7588001 budesonide with formoterol inh. 400mcg + 12mcg budesonide/formoterol
m8506001 SYMBICORT TURBOHALER 400mcg + 12mcg [ASTRAZENEC] budesonide/formoterol
m9605001 SYMBICORT TURBOHALER 100mcg + 6mcg [ASTRAZENEC] budesonide/formoterol
m9605002 SYMBICORT TURBOHALER 200mcg + 6mcg [ASTRAZENEC] budesonide/formoterol
m11694001 budesonide with formoterol inh. 100mcg + 6mcg budesonide/formoterol
m11694002 budesonide with formoterol inh. 200mcg + 6mcg budesonide/formoterol
m5374001 SERETIDE 250 EVOHALER 25mcg + 250mcg [A & H] salmeterol /fluticasone
m6016001 fluticasone with salmeterol inh. 50mcg + 25mcg salmeterol /fluticasone
m6016002 fluticasone with salmeterol inh. 125mcg + 25mcg salmeterol /fluticasone
m6016003 fluticasone with salmeterol inh. 250mcg + 25mcg salmeterol /fluticasone
m7800001 SERETIDE 500 ACCUHALER [GLAXO] salmeterol /fluticasone
m8452001 SERETIDE 125 EVOHALER 25mcg + 125mcg [A & H] salmeterol /fluticasone
m8651001 SERETIDE 250 ACCUHALER [GLAXO] salmeterol /fluticasone

Table A3b. Multilex codes for the drug exposures (*Continued*)

LABA:
m11474001 salmeterol with fluticasone inh. 50mcg + 100mcg salmeterol /fluticasone
m11474002 salmeterol with fluticasone inh. 50mcg + 250mcg salmeterol /fluticasone
m11474003 salmeterol with fluticasone inh. 50mcg+ 500mcg salmeterol /fluticasone
m11475001 fluticasone with salmeterol inh. 100mcg + 50mcg salmeterol /fluticasone
m11475002 fluticasone with salmeterol inh. 250mcg + 50mcg salmeterol /fluticasone
m11475003 fluticasone with salmeterol inh. 500mcg + 50mcg salmeterol /fluticasone
m11477001 SERETIDE 100 ACCUHALER [GLAXO] salmeterol /fluticasone
m12019001 salmeterol with fluticasone inh. 25mcg + 50mcg salmeterol /fluticasone
m12019002 salmeterol with fluticasone inh. 25mcg + 125mcg salmeterol /fluticasone
m12019003 salmeterol with fluticasone inh. 25mcg + 250mcg salmeterol /fluticasone
m12020001 SERETIDE 50 EVOHALER 25mcg + 50mcg [A & H] salmeterol /fluticasone
SABA (incl. combinations)
M2914001 salbutamol sulphate 100micrograms/inhalation Inhalation
M2914002 salbutamol sulphate 100micrograms/inhalation Inhalation
M1020001 salbutamol sulphate 100micrograms/inhalation Inhalation
M1193003 salbutamol sulphate Inhalation
M5056001 salbutamol sulphate 200micrograms Inhalation
M1020003 salbutamol sulphate 100micrograms/inhalation Inhalation
M5138001 salbutamol sulphate 100micrograms/actuation Inhalation
M8715001 salbutamol sulphate 100micrograms/actuation Inhalation
M11165001 salbutamol sulphate 100micrograms/actuation Inhalation
M11642001 salbutamol sulphate 95micrograms Inhalation
M1195001 salbutamol sulphate 100micrograms/actuation Inhalation
M5616001 salbutamol sulphate 100micrograms/actuation Inhalation
M11995001 salbutamol sulphate 100micrograms/actuation Inhalation
M5153001 salbutamol sulphate 200micrograms/blister Inhalation
M5153002 salbutamol sulphate 400micrograms/blister Inhalation
M1021002 salbutamol sulphate 400micrograms Inhalation
M8602001 salbutamol sulphate 100micrograms/inhalation Inhalation
M5056002 salbutamol sulphate 400micrograms Inhalation
M1021001 salbutamol sulphate 200micrograms Inhalation
M5616002 salbutamol sulphate 200micrograms/actuation Inhalation
M5544001 salbutamol sulphate 200micrograms Inhalation
M9581001 salbutamol 100micrograms/inhalation Inhalation
M1020002 salbutamol sulphate 200micrograms/actuation Inhalation
M10939001 salbutamol sulphate 100micrograms/actuation Inhalation

Table A3b. Multilex codes for the drug exposures (*Continued*)

SABA (incl. combinations)
M9412001 salbutamol sulphate 100micrograms/inhalation Inhalation
M8715002 salbutamol sulphate 100micrograms/actuation Inhalation
M8664001 salbutamol sulphate 100micrograms/actuation Inhalation
M5544002 salbutamol sulphate 400micrograms Inhalation
M8654001 salbutamol sulphate 100micrograms/inhalation Inhalation
M5616003 salbutamol sulphate 95micrograms Inhalation
M12722001 salbutamol sulphate 100micrograms/actuation Inhalation
M5800001 salbutamol sulphate 100micrograms/inhalation Inhalation
M11529001 salbutamol 100micrograms/inhalation Inhalation
M10652001 salbutamol sulphate 200micrograms/actuation Inhalation
M12723001 salbutamol sulphate 100micrograms/actuation Inhalation
M11737001 salbutamol sulphate 100micrograms/inhalation Inhalation
M9580001 salbutamol 100micrograms/inhalation Inhalation
M12724001 salbutamol sulphate 200micrograms/actuation Inhalation
M2414001 salbutamol sulphate 100micrograms Inhalation
M10724001 salbutamol 100micrograms/inhalation Inhalation
M1561009 salbutamol sulphate 100micrograms/inhalation Inhalation
M3371009 salbutamol sulphate 100micrograms/inhalation Inhalation
M1340001 salbutamol sulphate 200micrograms Inhalation
M8727001 salbutamol sulphate 100micrograms/actuation Inhalation
M5138002 salbutamol sulphate 100micrograms/actuation Inhalation
M606009 salbutamol sulphate 100micrograms/inhalation Inhalation
M3974009 salbutamol sulphate 400micrograms Inhalation
M3834009 salbutamol sulphate 100micrograms/inhalation Inhalation
M3973009 salbutamol sulphate 200micrograms Inhalation
M3164009 salbutamol sulphate 100micrograms/inhalation Inhalation
M3897009 salbutamol sulphate 100micrograms/inhalation Inhalation
M1340002 salbutamol sulphate 400micrograms Inhalation
M5137001 salbutamol sulphate 100micrograms/actuation Inhalation
M3435009 salbutamol sulphate 100micrograms/inhalation Inhalation
M602009 salbutamol sulphate 100micrograms/inhalation Inhalation
M3574009 salbutamol sulphate 100micrograms/inhalation Inhalation
M603009 salbutamol sulphate 100micrograms/inhalation Inhalation
M16041001 salbutamol sulphate 100micrograms Inhalation
M2032009 salbutamol sulphate 200micrograms Inhalation
M16043001 salbutamol sulphate 100micrograms Inhalation

Table A3b. Multilex codes for the drug exposures (*Continued*)

SABA (incl. combinations)
M16042001 salbutamol sulphate 100micrograms Inhalation
M16044001 salbutamol sulphate 100micrograms Inhalation
M2032010 salbutamol sulphate 400micrograms Inhalation
M5470001 salbutamol sulphate 100micrograms/actuation Inhalation
M1023001 salbutamol sulphate 5mg/ml Nebulised
M1024001 salbutamol sulphate 2.5mg Nebulised
M2221002 salbutamol sulphate 5mg/2.5ml Nebulised
M5055001 salbutamol sulphate 2.5mg/2.5ml Nebulised
M5055002 salbutamol sulphate 5mg/2.5ml Nebulised
M1024002 salbutamol sulphate 5mg Nebulised
M2221001 salbutamol sulphate 2.5mg/2.5ml Nebulised
M10821001 salbutamol sulphate 5mg/2.5ml Nebulised
M8679001 salbutamol sulphate 2.5mg/2.5ml Nebulised
M5056003 salbutamol sulphate 5mg/ml Nebulised
M8429001 salbutamol sulphate 2.5mg/2.5ml Nebulised
M8429002 salbutamol sulphate 5mg/2.5ml Nebulised
M5740001 salbutamol sulphate 5mg/ml Nebulised
M2901010 salbutamol sulphate 5mg/2.5ml Nebulised
M2901009 salbutamol sulphate 2.5mg/2.5ml Nebulised
M16669001 salbutamol sulphate 5mg/2.5ml Nebulised
M6698009 salbutamol sulphate 2.5mg/2.5ml Nebulised
M1340003 salbutamol sulphate Unknown
M6267001 salbutamol/sodium cromoglicate Inhalation
M6268001 salbutamol/sodium cromoglicate Inhalation
M6268002 salbutamol/sodium cromoglicate Inhalation
M6267002 salbutamol/sodium cromoglicate Inhalation
M7243001 ipratropium bromide/salbutamol sulphate 20mcg + 100mcg Inhalation
M7244001 ipratropium bromide/salbutamol sulphate 20mcg + 100mcg Inhalation
M8688001 ipratropium bromide/salbutamol sulphate 2.5ml Nebulised
M8689001 ipratropium bromide/salbutamol sulphate 500micrograms + 2.5mg/2.5ml Nebulised
M8892001 ipratropium bromide/salbutamol sulphate 2.5mg + 500micrograms/2.5ml Unknown
M9319001 ipratropium bromide/salbutamol sulphate 100micrograms + 20micrograms/actuation Unknown
M14551001 ipratropium bromide/salbutamol sulphate 500micrograms + 2.5mg/2.5ml Nebulised
M106003 fenoterol hydrobromide 100micrograms/actuation Inhalation
M106001 fenoterol hydrobromide 200micrograms/actuation Inhalation

Table A3b. Multilex codes for the drug exposures (*Continued*)

SABA (incl. combinations)
M106002 fenoterol hydrobromide 5mg/ml Nebulised
M3445003 fenoterol hydrobromide 100micrograms/actuation Inhalation
M3445001 fenoterol hydrobromide 200micrograms/actuation Inhalation
M281001 fenoterol hydrobromide/ipratropium bromide 40micrograms + 100micrograms/actuation Inhalation
M745001 fenoterol hydrobromide/ipratropium bromide 500micrograms + 1.25mg/4ml Nebulised
M743001 fenoterol hydrobromide/ipratropium bromide Nebulised
M7105001 fenoterol hydrobromide/ipratropium bromide Inhalation
M3446001 fenoterol hydrobromide/ipratropium bromide 100micrograms + 40micrograms/actuation Inhalation
M3446002 fenoterol hydrobromide/ipratropium bromide 100micrograms + 40micrograms/actuation Inhalation
M8921001 fenoterol hydrobromide/ipratropium bromide 1.25mg + 500micrograms/4ml Unknown
M3889001 fenoterol hydrobromide/ipratropium bromide 40micrograms + 100micrograms/actuation Unknown
M3889002 fenoterol hydrobromide/ipratropium bromide 40micrograms + 100micrograms/actuation Unknown
M121002 terbutaline sulphate 500micrograms Inhalation
M6845001 terbutaline sulphate 500micrograms Inhalation
M4785001 terbutaline sulphate 250micrograms/actuation Inhalation
M4785003 terbutaline sulphate 250micrograms/actuation Inhalation
M2757001 terbutaline sulphate Inhalation
M4785002 terbutaline sulphate 250micrograms/actuation Inhalation
M121003 terbutaline sulphate Inhalation
M122002 terbutaline sulphate 10mg/ml Nebulised
M122001 terbutaline sulphate 5mg/2ml Nebulised
M4786001 terbutaline sulphate 5mg/2ml Nebulised
M4786002 terbutaline sulphate 10mg/ml Nebulised
M2822009 terbutaline sulphate 2.5mg/ml Nebulised
M121001 terbutaline sulphate Inhalation
M6302001 orciprenaline sulphate 750micrograms/inhalation Inhalation
M33001 orciprenaline sulphate 750micrograms/inhalation Inhalation
M6302002 orciprenaline sulphate 750micrograms/inhalation Inhalation
M32003 orciprenaline sulphate 750micrograms/inhalation Inhalation
M1439001 pirbuterol acetate Inhalation
M6294001 pirbuterol acetate Inhalation
m4016001 beclometasone/salbutamol aerosol inh 50mcg/100mcg/inh beclometasone/salbutamol
m4016003 beclometasone/salbutamol caps (for inh) 100mcg/200mcg beclometasone/salbutamol

Table A3b. Multilex codes for the drug exposures (*Continued*)

SABA (incl. combinations)
m4016002 beclometasone/salbutamol caps (for inh) 200mcg/400mcg beclometasone/salbutamol
m5058001 salbutamol/beclometasone aerosol inh 100mcg/50mcg beclometasone/salbutamol
m5058003 salbutamol/beclometasone caps (for inh) 200mcg/100mcg beclometasone/salbutamol
m5058002 salbutamol/beclometasone caps (for inh) 400mcg/200mcg beclometasone/salbutamol
m1419001 VENTIDE aerosol inh [A & H] beclometasone/salbutamol
m5150002 VENTIDE paediatric ROTACAPS [A & H] beclometasone/salbutamol
m17182001 SALIPRANEB neb sol 500mcg + 2.5mg/2.5ml ipratropium/salbutamol
m128001 BRONCHODIL aerosol inh 500mcg/dose reproterol
m4672001 reproterol aerosol inh 500mcg/dose reproterol
m4672002 reproterol respirator sol 10mg/ml reproterol
m801001 PULMADIL aerosol inh [3M] rimiterol
m1192001 PULMADIL AUTO aerosol inh [3M] rimiterol
m4689001 rimiterol aerosol inh rimiterol
m605009 SALBUTAMOL aerosol inh 100mcg/inh salbutamol
m7949009 SALBUTAMOL cfc free inh 100mcg/inh [NEOLAB] salbutamol
m3164011 SALBUTAMOL neb sol 5mg/2.5ml [GEN (UK)] salbutamol
m16667001 STERIPOULE SALBUTAMOL neb sol 2.5mg/2.5ml salbutamol
m17740001 VENTOLIN ACCUHALER dry powder inh 200mcg salbutamol
m17739001 VENTOLIN EVOHALER cfc free inh 100mcg/inh salbutamol
m122002 BRICANYL respirator sol 10mg/ml [ASTRAZENEC] terbutaline
m17804001 BRICANYL RESPULES neb sol 5mg/2ml terbutaline
m17738001 BRICANYL TURBOHALER dry powder inh 500mcg terbutaline
SAMA
M69001 ipratropium bromide 20micrograms/actuation Inhalation
M1722001 ipratropium bromide 0.25mg/ml Nebulised
M2177009 ipratropium bromide 250micrograms/ml Nebulised
M2665001 ipratropium bromide 40micrograms/actuation Inhalation
M3114009 ipratropium bromide 250micrograms/ml Nebulised
M3886001 ipratropium bromide 20micrograms/dose Inhalation
M3886002 ipratropium bromide 20micrograms/dose Inhalation
M3887001 ipratropium bromide 40micrograms/metered inhalation Inhalation
M3888001 ipratropium bromide 0.25mg/ml Nebulised
M3888002 ipratropium bromide 250micrograms/ml Nebulised
M3888003 ipratropium bromide 250micrograms/ml Nebulised
M5411009 ipratropium bromide 250micrograms/ml Nebulised
M6314001 ipratropium bromide 40mcg Inhalation

Table A3b. Multilex codes for the drug exposures (*Continued*)

SAMA
M6314002 ipratropium bromide 40mcg Inhalation
M6315001 ipratropium bromide 40mcg Inhalation
M6315002 ipratropium bromide 40mcg Inhalation
M7104001 ipratropium bromide 20micrograms/actuation Inhalation
M7334001 ipratropium bromide 250micrograms/ml Nebulised
M11700001 ipratropium bromide 250micrograms/ml Nebulised
M11726001 ipratropium bromide 250micrograms/ml Nebulised
M12485001 ipratropium bromide 20micrograms/actuation Inhalation
M12486001 ipratropium bromide 20micrograms/actuation Inhalation
M12871001 ipratropium bromide 250micrograms/ml Nebulised
M12872001 ipratropium bromide 500micrograms/2ml Nebulised
M12873001 ipratropium bromide 250micrograms/1ml Nebulised
M12874001 ipratropium bromide 500micrograms/2ml Nebulised
M12890001 ipratropium bromide 250micrograms/ml Nebulised
M12891001 ipratropium bromide 500micrograms/2ml Nebulised
M12892001 ipratropium bromide 250micrograms/ml Nebulised
M12894001 ipratropium bromide 500micrograms/2ml Nebulised
M16659001 ipratropium bromide 250micrograms/ml Nebulised
M16663001 ipratropium bromide 500micrograms/2ml Nebulised
m17737001 ATROVENT AEROCAPS inh powder caps 40mcg ipratropium
m17736001 ATROVENT AEROHALER inh powder caps/device 40mcg ipratropium
LAMA
M6831001 oxitropium bromide 100micrograms/actuation Inhalation
M6832001 oxitropium bromide 100micrograms/actuation Inhalation
M6832002 oxitropium bromide 100micrograms/actuation Inhalation
M7103001 oxitropium bromide 100micrograms/actuation Inhalation
M6542001 tiotropium bromide monohydrate 18 micrograms Inhalation
M10764001 tiotropium bromide monohydrate 18 micrograms Inhalation
M14945001 tiotropium bromide monohydrate 18 micrograms Inhalation
M14946001 tiotropium bromide monohydrate 18 micrograms Inhalation
M14947001 tiotropium bromide monohydrate 18 micrograms Inhalation
M14948001 tiotropium bromide monohydrate 18 micrograms Inhalation
M15641001 tiotropium bromide monohydrate 2.5 micrograms/actuation Inhalation
M15642001 tiotropium bromide monohydrate 2.5 micrograms/actuation Inhalation

LABA: long acting beta2-agonist; SABA: short acting beta2-agonist; SAMA: short-acting muscarinic antagonist; LAMA: long-acting muscarinic antagonist.

Table A4a. International Classification of Primary Care codes for outcome of interest

K75	Myocardial Infarction

Table A4b. Read Codes lists for outcome of interest

Myocardial infarction (incident)
323..00 ecg: myocardial infarction
3233..00 ecg: antero-septal infarct.
3234..00 ecg:posterior/inferior infarct
3235..00 ecg: subendocardial infarct
3236..00 ecg: lateral infarction
323z.00 ecg: myocardial infarct nos
G30..00 acute myocardial infarction
G30..11 attack - heart
G30..12 coronary thrombosis
G30..13 cardiac rupture following myocardial infarction (mi)
G30..14 heart attack
G30..15 mi - acute myocardial infarction
G30..16 thrombosis - coronary
G30..17 silent myocardial infarction
G300.00 acute anterolateral infarction
G301.00 other specified anterior myocardial infarction
G301000 acute anteroapical infarction
G301100 acute anteroseptal infarction
G301z00 anterior myocardial infarction nos
G302.00 acute inferolateral infarction
G303.00 acute inferoposterior infarction
G304.00 posterior myocardial infarction nos
G305.00 lateral myocardial infarction nos
G306.00 true posterior myocardial infarction
G307.00 acute subendocardial infarction
G307000 acute non-q wave infarction
G307100 acute non-st segment elevation myocardial infarction
G308.00 inferior myocardial infarction nos
G309.00 acute q-wave infarct
G30a.00 mural thrombosis
G30b.00 acute posterolateral myocardial infarction

Table A4b. Read Codes lists for outcome of interest (*Continued*)

Myocardial infarction (incident)
G30x.00 acute transmural myocardial infarction of unspecif site
G30x000 acute st segment elevation myocardial infarction
G30y.00 other acute myocardial infarction
G30y000 acute atrial infarction
G30y100 acute papillary muscle infarction
G30y200 acute septal infarction
G30yz00 other acute myocardial infarction nos
G30z.00 acute myocardial infarction nos
G310.00 postmyocardial infarction syndrome
G31y100 microinfarction of heart
G35.00 subsequent myocardial infarction
G350.00 subsequent myocardial infarction of anterior wall
G351.00 subsequent myocardial infarction of inferior wall
G353.00 subsequent myocardial infarction of other sites
G35x.00 subsequent myocardial infarction of unspecified site
G36..00 certain current complication follow acute myocardial infarct
G360.00 haemopericardium/current comp folow acut myocard infarct
G361.00 atrial septal defect/curr comp folow acut myocardal infarct
G362.00 ventric septal defect/curr comp fol acut myocardal infarctn
G363.00 ruptur cardiac wall w'out haemopericard/cur comp fol ac mi
G364.00 ruptur chordae tendinae/curr comp fol acute myocard infarct
G365.00 rupture papillary muscle/curr comp fol acute myocard infarct
G366.00 thrombosis atrium,auric append&vent/curr comp foll acute mi
G38.00 postoperative myocardial infarction
G380.00 postoperative transmural myocardial infarction anterior wall
G381.00 postoperative transmural myocardial infarction inferior wall
G382.00 postoperative transmural myocardial infarction other sites
G383.00 postoperative transmural myocardial infarction unspec site
G384.00 postoperative subendocardial myocardial infarction
G38z.00 postoperative myocardial infarction, unspecified
Gyu3100 [x]other current complicatns following acute myocard infarct
Gyu3400 [x]acute transmural myocardial infarction of unspecif site
Gyu3500 [x]subsequent myocardial infarction of other sites
Gyu3600 [x]subsequent myocardial infarction of unspecified site

CHAPTER 4.2

Instrumental Variables Analysis Using Multiple Databases: an Example of Antidepressant Use and Risk of Hip Fracture

Md Jamal Uddin, Rolf HH Groenwold, Anthonius de Boer, Helga Gardarsdottir, Elisa Martin, Gianmario Candore, Svetlana V Belitser, Arno W Hoes, Kit CB Roes, Olaf H Klungel

Submitted, Pharmacoepidemiology and Drug Safety

ABSTRACT

BACKGROUND: Instrumental variable (IV) analysis can control for unmeasured confounding, yet it has not been widely used in pharmacoepidemiology. We aimed to assess the performance of IV analysis using different IVs in multiple databases in a study of antidepressant use and hip fracture.

METHODS: Information on adults with at least one prescription of a selective serotonin reuptake inhibitor (SSRI) or tricyclic antidepressant (TCA) during 2001-2009 was extracted from the THIN (UK), BIFAP (Spain), and Mondriaan (Netherlands) databases. IVs were created using the proportion of SSRI prescriptions per practice or using the one, five, or ten previous prescriptions by a physician. Data were analysed using conventional Cox regression and two-stage IV models.

RESULTS: In the conventional analysis, SSRI (vs. TCA) was associated with an increased risk of hip fracture, which was consistently found across databases: the adjusted hazard ratio (HR) was approximately 1.35 for time-fixed and 1.50 to 2.49 for time-varying SSRI use, while the IV analysis based on the IVs that appeared to satisfy the IV assumptions showed conflicting results, e.g., the adjusted HRs ranged from 0.55 to 2.75 for time-fixed exposure. All IVs for time-varying exposure violated at least one IV assumption and were therefore invalid.

CONCLUSIONS: Our multiple database study shows that the performance of IV analysis varied across the databases for time-fixed and time-varying exposures and strongly depends on the definition of IVs. It remains challenging to obtain valid IVs in pharmacoepidemiological studies, particularly for time-varying exposure, and IV analysis should therefore be interpreted cautiously.

INTRODUCTION

Electronic healthcare databases are being used to detect unintended and intended effects of drugs in comparative safety and effectiveness research. However, such databases sometimes have very limited or inaccurate information on potential confounding variables, such as alcohol consumption and functional health status (e.g., activities of daily living),[1,2] which may impair the validity of study results. For example, several observational studies[3–8] indicated an association between antidepressants use (mainly tricyclic antidepressants and selective serotonin re-uptake inhibitors) and the risk of hip fracture. However, results from these studies are heterogeneous for many reasons.[3,5,6,8] Among them, confounding by unmeasured patients characteristics such as severity of depression and lifestyle factors (e.g., smoking) have usually not been accounted for in most of the studies.[3] Instrumental variable (IV) analysis has been used to control for unmeasured confounding of comparative safety and effectiveness studies in pharmacoepidemiology.[2,9–15]

An IV is a variable that can be considered to mimic the treatment allocation process in a randomized study.[2] That means an IV 1) is associated with the exposure, 2) affects the outcome only through the exposure, and 3) is independent of confounders.[16–18] If these key assumptions and some additional assumptions (e.g., monotonicity) are satisfied, IV analysis may consistently estimate the average causal effect of an exposure on an outcome.[18,19]

Several studies[10,11,13–15,20–29] have applied IV analysis using electronic healthcare databases on various pharmacoepidemiological issues, for example selective cyclooxygenase-2 inhibitors and upper gastrointestinal complications and myocardial infarction.[13] As far as we know, there is no study of IV analysis that assessed the performance of IV analysis across multiple databases in different countries using both time-fixed and time-varying exposures. We therefore aimed to assess the performance of IV analysis using different IVs in general practice (GP) databases from three European countries (Spain, UK, and Netherlands) in a study of antidepressants use and risk of hip/femur fracture.

METHODS

Data sources and study population

This study was conducted using data from the health improvement network, UK (THIN); the Spanish Base de datos para la Investigación Farmacoepidemiológica en Atención Primaria (BIFAP); and the Dutch Mondriaan GP database.[30–32] The detailed information of databases and data specifications can be found in a common study protocol, which is available online.[33]

We sampled 587,637 patients from THIN, 252,203 patients from BIFAP, and 22,954 patients from Mondriaan who were aged ≥18 years and had at least one prescription of selective serotonin re-uptake inhibitors (SSRI) or tricyclic antidepressants (TCA) during the study period (1 January 2001 to 31 December 2009). Moreover, at least one year of enrolment in the database without any antidepressants prescription in the six months and 1 year without a hip fracture preceding the enrolment in the cohort were required to enter the study. The index date (date of cohort entry) is defined for each patient as the date of first SSRI or TCA prescription within the study period. The observation period for each patient lasts from the index date to the end of data collection (31 December 2009), the date of the first fracture, the date of death, or loss to follow-up, whichever occurs first. We also excluded patients (17,498 in THIN, 1,338 in BIFAP and 480 in Mondriaan) from the study population because of combined use of SSRI and TCA or subsequent use within the same treatment episode and all subjects from practices that contributed less than 50 patients; the latter because a considerable number of prescriptions is needed to accurately measure the IV.[20] The final number of practices was 502 for THIN, 280 for BIFAP, and 133 for Mondriaan.

Exposures and outcome

We considered two types of exposure: 1) exposure that was determined at baseline/on index date and considered constant over time; and 2) exposure that was time-varying.

The first type of exposure was defined based on the first SSRI (vs. TCA) prescription, and consequently the follow-up time was the time between the first prescription and the end of follow-up. During the follow-up, we ignored any changes in medication status (e.g., switch from TCA to SSRI or stopping SSRI use). Follow-up time for time-varying exposure was divided into periods of "current use" to SSRI/TCA (from the beginning of the prescription up to the calculated end date of the last prescription supply in a continuous treatment episode, i.e. < 30 days between the end of a prescription supply and the beginning of the next one), "recent use" (during the 91 days following the calculated end date of the prescription/current use/continuous treatment period), and "past use" (after 91 days following the calculated end of prescription).[34] A patient can switch between current, recent and past periods and between the treatment classes (e.g., SSRI to TCA). The choice for the 91days in the calculation interval was based on the fact that Dutch health insurance policies cover the dispensing of the majority of drugs for three months.[35,36] The period of "past use" ends when the patient becomes a new user or at the end of follow-up. If a patient switches between the treatment classes, a new treatment period starts at the date of the prescription of the new drug (SSRI/TCA). For the second type of exposure, antidepressant use was re-categorised into two groups: current SSRI users and non-SSRI users. Non-SSRI users could be current, recent or past TCA users or recent/past SSRI users. Throughout the manuscript we will use the abbreviations "$SSRI_{fixed}$" and "$SSRI_{time}$" for the exposure that was considered constant during follow-up and time-varying exposure, respectively. The drug codes are provided in the study protocol.[33]

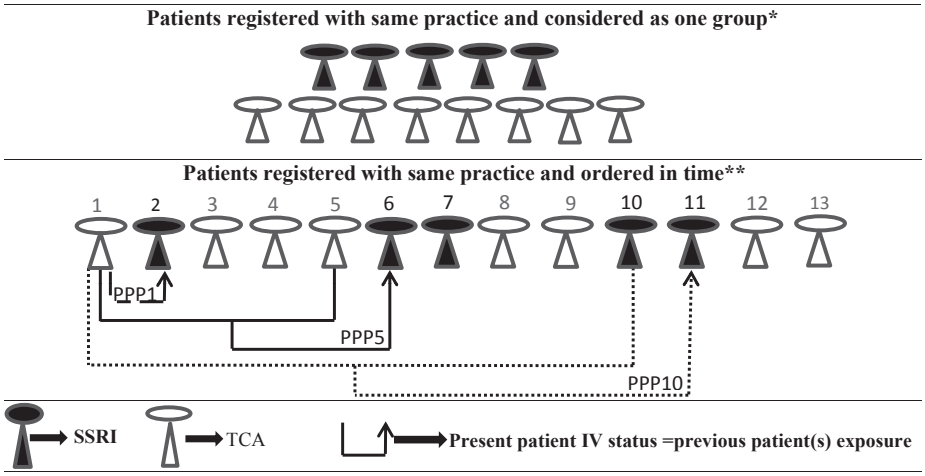
A first fracture of the hip/femur (HF) during the study period regardless of whether they have a history of past fractures was the outcome of interest. In BIFAP, HF of patients who, after the review of their automated clinical records, are shown to be a result of major trauma (e.g., car accident one month before) were excluded. A detailed description of the outcome is provided in the study protocol.[33]

Potential confounding factors

The potential confounding factors (i.e., co-medications, co-morbidities, and lifestyle factors) are listed in the Table 1. For the exposure $SSRI_{fixed}$, confounders were assessed at baseline and for the exposure $SSRI_{time}$ confounders were considered time-varying and assessed at baseline and updated whenever patients switched between exposure status (current, recent, past) or every six months (if exposure status did not change). For co-morbidities, patients were classified as having the disease from the first date of diagnosis onwards. More details on confounding factors are available online.[33]

Construction of instrumental variables

We considered several IVs based on the physician's prescribing preferences (PPPs) of SSRI or TCA (Figure 1).[2,20] For $SSRI_{fixed}$, the PPPs were based on either the single last prescription (PPP1), the last five consecutive prescriptions (PPP5), or the last ten consecutive prescriptions (PPP10). Furthermore, we used the proportion of SSRI prescriptions (PSP) per practice



Patients registered with same practice	Actual Prescription	PSP*	PPP1**	PPP5**	PPP10**
Patient 1	TCA	5/13	-	-	-
Patient 2	**SSRI**	5/13	0	-	-
Patient 3	TCA	5/13	1	-	-
Patient 4	TCA	5/13	0	-	-
Patient 5	TCA	5/13	0	-	-
Patient 6	**SSRI**	5/13	0	1/5	-
Patient 7	**SSRI**	5/13	1	2/5	-
Patient 8	TCA	5/13	1	2/5	-
Patient 9	TCA	5/13	0	2/5	-
Patient 10	**SSRI**	5/13	0	2/5	-
Patient 11	**SSRI**	5/13	1	3/5	4/10
Patient 12	TCA	5/13	1	3/5	5/10
Patient 13	TCA	5/13	0	2/5	4/10

Time-varying IV

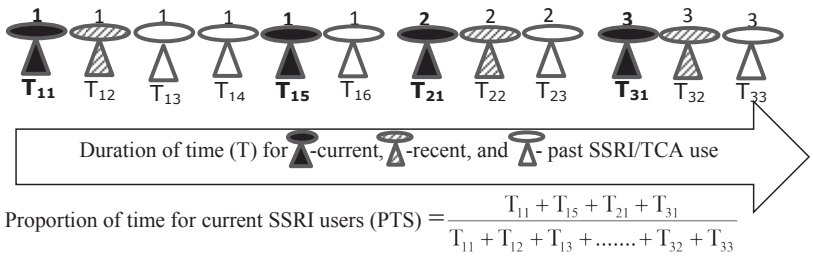


Figure 1. Definition of instrumental variables based on proportion of SSRI prescriptions and previous prescription written by physician.

SSRI: Selective serotonin re-uptake inhibitor, TCA: Tricyclic antidepressant, PSP: proportion of SSRI prescription, PPP1: physician prescribing preference with single last prescription (binary, 1=SSRI or 0=TCA); PPP5: physician prescribing preference with previous five prescriptions, e.g., when patients are ordered in time, IV for patient number 6 is 1/5 (i.e., among the last 5 prescriptions of patient 6, 1 prescriptions are SSRI), PPP10- physician prescribing preference with previous ten prescriptions, e.g., when patients are ordered in time, IV for patient number 11 is 4/10 (i.e., among the last 10 prescriptions of patient 11, 4 prescriptions are SSRI). PPP5/PPP10 is continuous variable (0-100%) *All subjects under a practice are considered as one group; **Subjects are ordered in time by practice.

to create an alternative preference based IV. We dichotomized the PSP at its median to create a binary IV (PSP_{dich}). Finally, we also considered other IVs based on the last 20 or 50 prescriptions issued by physicians within a practice.

For the $SSRI_{time}$, the IVs were measured using the proportions of time of "current" SSRI use (PTS) in a practice (Figure 1). Specifically, the PTS is the ratio of the follow-up time contributed by patients who are current users of SSRI under a practice to the total follow-up time contributed by patients under the same practice. We also dichotomized the PTS (at the median) to create a binary IV (PTS_{dich}).

Assessment of instrumental variable assumptions

IV assumption 1: We used point bi-serial correlation (r) for binary exposure and continuous IV and odds ratio (OR) for both binary exposure and IV to check the association between IV and exposure.[11,18] Although the partial F-statistic value from the first-stage IV model is mostly used to assess this assumption, we did not use it because it is highly influenced by sample size.[11,18]

IV assumptions 2 and 3: We assumed that a physician's preference did not influence the risk of HF of a patient (assumption 2) nor was it correlated with confounders (assumption 3).[20] In addition, a falsification test was applied to assess assumption 3 using the standardized difference (SDif) to assess balance of measured confounders between IV categories (when an IV was continuous, balance was assessed across the quintiles).[37] We also measured the balance on several confounders simultaneously using a multivariate distance measure, Mahalanobis distance (MD).[23,38] Lower values of SDif (e.g., < 0.10)[39] and MD indicate better balance.[37,38] Once measured confounders are insufficiently balanced between IV categories, this may also imply imbalance of unmeasured confounders (the assumption 3 is violated), and thus IV analysis is inappropriate.[37] Nevertheless, if measured confounders are balanced, we assumed that such balance could be carried-over to unmeasured confounders.[2,27]

Statistical analyses

We analysed the three datasets separately. We also analysed the pooled data for the exposure $SSRI_{fixed}$. A conventional Cox proportional hazards model and two-stage IV models were used to analyse the data. The first-stage model of the IV analysis was a linear regression model,[40,41] where the exposure ($SSRI_{fixed}$ or $SSRI_{time}$) was the dependent variable and the IV was the independent variable. The second-stage model was a Cox proportional hazard model, in which predicted treatment rather than actual treatment was the independent variable.[18] We estimated the unadjusted and adjusted hazard ratios without and with including confounders (baseline confounders for the $SSRI_{fixed}$ and the time-varying confounders for $SSRI_{time}$) in both conventional and IV analyses. Bootstrapping (1000 bootstrap samples) was applied to estimate the confidence intervals (CI) in our IV analyses. All analyses were performed using the statistical software R version 2.15.2.[42]

The effects of SSRI/TCA on the risk of HF may not be homogeneous in our study population. Hence, the point estimate of average causal effect of exposure for the whole study population was not possible to identify and other point estimates (e.g., complier average causal effects (CACE) or local average treatment effect (LATE)) can be identified.[16,18] To identify the CACE, a fourth assumption, monotonicity (the IV affects the exposure deterministically in one direction i.e. there are no defiers) was considered.[16,19,43,44] In our example, the compliers are subjects who would be prescribed a SSRI had they seen a physician who preferred SSRI, but would be prescribed a TCA had they seen a physician who preferred TCA.[13,19] For continuous IVs, the CACE is a weighted average of the effect in multiple subgroups where the more compliant subjects obtain a larger weight.[16,19,43,44]

RESULTS

Table 1 shows the patients characteristics, stratified by exposure status (SSRI vs. TCA) at the index date and database (n = 570139 for THIN, n = 252203 for BIFAP, and n = 22474 for Mondriaan). In total, 3631 persons in THIN and 1528 persons in BIFAP and 81 persons in Mondriaan experienced a HF during the study period. As expected, measured confounders were more balanced between IV categories than between exposure groups (Table A2a-A2b, Appendix).

Table 1. Characteristics of patients stratified by database and type of antidepressant drugs (SSRI vs. TCA) assigned at index date

	THIN		BIFAP		Mondriaan	
	TCA	SSRI	TCA	SSRI	TCA	SSRI
Sample size	220562	349577	44599	207604	8033	14441
Number of cases (HF)	1694	1937	240	1288	32	49
Confounders (%)						
Mean age (SD)	56.0 (17.5)	46.0 (18.06)	53.0 (16.8)	50.0 (17.8)	54.0 (16.9)	46.0 (16.8)
Sex	37.2	35.6	25.9	27.6	35.9	36.6
Stroke	4.7	3.5	3.6	3.8	2.7	2.1
Ischaemic heart disease	9.1	5.6	4.6	4.9	5.6	3.8
Dementia	0.3	0.9	0.6	1.3	-	-
History of fractures	18.9	18.5	4.7	4.6	5.5	4.4
Mental disorders	1.7	2.8	1.5	2.2	2.7	4.8
Osteoporosis	3.7	1.8	6.5	5.1	2.7	1.3
Anaemia	7.8	6.5	7.7	7.3	5.0	4.4
DMARD	1.6	0.7	0.8	0.5	1.7	0.7

Table 1. Characteristics of patients stratified by database and type of antidepressant drugs (SSRI vs. TCA) assigned at index date (*Continued*)

	THIN		BIFAP		Mondriaan	
	TCA	SSRI	TCA	SSRI	TCA	SSRI
Benzodiazepines	14.1	17.5	29.4	41.8	29.3	37.0
Antiepileptic drugs	4.0	2.0	7.1	3.6	5.9	1.6
Antidiabetic drugs	6.4	3.5	5.9	5.3	8.6	4.2
Glucocorticoids	6.3	3.7	3.5	2.6	7.1	3.3
Antihypertensive	29.2	18.9	20.1	18.6	24.7	16.1
ACE inhibitors	-	-	9.0	8.4	8.7	5.7
Angiotensin II antagonists	-	-	4.9	5.1	5.4	2.6
Diuretics	17.8	10.3	9.8	9.1	13.3	7.5
Opioids (including Morphine)	19.3	6.2	13.1	5.4	16.8	4.1
Hormone replacement therapy	7.2	5.5	1.9	1.5	3.5	3.0
Other Antidepressant	1.9	2.9	3.0	3.5	2.2	3.9
Antipsychotics	4.6	5.2	4.8	5.6	1.7	3.1
Vitamin-D	0.3	0.1	5.3	3.9	2.4	1.2
Antiparkinson	0.6	0.6	0.7	0.9	0.8	0.6
Statins	15.7	8.9	9.9	9.1	12.3	7.3
Other information						
SSRI users (%)	61.3		82.2		64.0	
Total follow-up time in years	2118320		757403		63806	
Median follow-up time in years	3.37		2.71		2.25	
Number of Practices	502		280		133	

SSRI- Selective serotonin re-uptake inhibitor; TCA- Tricyclic antidepressant; THIN- The Health Improvement Network, UK; BIFAP- the Spanish Base de datos para la Investigación Farmacoepidemiológica en Atención Primaria; Mondriaan- Dutch general practices databases.

Several methods were applied to evaluate the validity of IVs particularly the association between exposure and IV (assumption 1) and independence between IV and confounders (assumption 3). The summary results are presented in Table 2. In THIN, the IVs, PSP and PPP10, were strongly associated with $SSRI_{fixed}$ ($r \geq 0.15$, Table A1, Appendix) and confounders were balanced across the quintiles of these IVs (SDif < 0.10, Table A2a, Appendix). Similar performance was observed for the PSP in BIFAP and PPP5 and PPP10 in Mondriaan. In many situations in which an IV was weakly associated with the exposure, the confounders were balanced with that IV or vice versa. For instance, the IV PPP1 was weakly associated with the $SSRI_{fixed}$ (OR < 2 in all databases, Table A1, Appendix), but the measured confounders were balanced across the categories of PPP1 (SDif < 0.10 in all databases, Table A2a-b, Appendix).

Table 2. Summary results from instrumental variables analysis

Database	Instrumental Variables	IV Assumption 1	IV Assumption 3	Estimates Unadj-Adj	Length CI: Adj-Model
THIN	**PSP**			**0.51-1.09**	**0.85**
	PSP_{dich}		X	0.52-1.07	1.29
	PPP1	X		1.00-2.02	5.74
	PPP5	X		0.62-1.31	1.75
	PPP10			**0.57-1.16**	**1.22**
	PTS*	X		0.97-1.90	4.50
	PTS_{dich}*	X		0.50-0.97	4.34
BIFAP	**PSP**			**4.51-2.75**	**6.13**
	PSP_{dich}	X		3.18-1.86	7.10
	PPP1	X		22.2-42.2	5258
	PPP5	X		3.07-3.44	26.6
	PPP10	X		2.57-1.89	9.02
	PTS*	X		79.0-4.04	15.6
	PTS_{dich}*	X	X	45.8-4.21	46.5
Modriaan	PSP		X	0.28-1.19	8.05
	PSP_{dich}		X	0.24-1.10	15.3
	PPP1	X		0.09-0.20	93.5
	PPP5			**0.20-0.55**	**16.2**
	PPP10			**0.44-1.67**	**27.5**
	PTS*	X	X	0.44-0.83	25.8
	PTS_{dich}*	X	X	0.03-0.27	32.4

SSRI- Selective serotonin re-uptake inhibitor; TCA- Tricyclic antidepressant; THIN- The Health Improvement Network, UK; BIFAP- the Spanish Base de datos para la Investigación Farmacoepidemiológica en Atención Primaria; Mondriaan- Dutch general practices databases; IV- instrumental variable; PSP- proportion of SSRI prescriptions per practice; PSP_{dich}- proportion of SSRI prescriptions per practice (dichotomized at the median); PPP1- single previous prescription by a physician; PPP5- previous five prescriptions by a physician; PPP10- previous ten prescriptions by a physician; PTS- proportion of time of SSRI prescriptions; PTS_{dich}- proportion of time of SSRI prescriptions (dichotomized at the median); "X" indicates instrumental variables assumptions are violated; Unadj- estimates from IV models without confounders; Adj- estimates from IV models that included confounders; CI-confidence interval; **Bold** numbers indicate the potential instrumental variables that at least satisfied assumption 1 and 3;*IVs for time-varying settings.

Alternatively, PSP and PSP_{dich} were strongly associated with $SSRI_{fixed}$ (r = 0.26 and OR = 2.44, respectively, Table A1) in Mondriaan, but confounders were imbalanced (e.g., difference in mean age, Table A2a, Appendix). All IVs, except PSP in BIFAP, were weakly associated with the $SSRI_{fixed}$ and exposure effect estimates based on the weak IVs had very wide confidence intervals (Table 3). When the number of previous prescriptions increased for estimating the IV PPP (e.g., previous 20 or 50 prescriptions), the IV estimates were approximately similar to

Table 3. Associations between SSRI vs. TCA use and risk of hip fracture based on conventional and IV Analysis

		Conventional Cox Model	Instrumental Variable analysis				
			PSP	PSP_{dich}	PPP1	PPP5	PPP10
Database	Model	HR [CI]	HR [CI]*	HR [CI] *	HR [CI] *	HR [CI] *	HR [CI] *
THIN	Unadjusted	0.72 [0.67-0.77]	**0.51 [0.36-0.70]**	0.51 [0.31-0.80]	1.00 [0.39-2.69]	0.62 [0.37-1.08]	**0.57 [0.36-0.92]**
	Adjusted	1.35 [1.26-1.44]	**1.09 [0.75-1.60]**	1.23 [0.72-2.01]	2.02 [0.61- 6.35]	1.31 [0.73- 2.48]	**1.16 [0.70-1.92]**
BIFAP	Unadjusted	1.21 [1.06-1.39]	**4.51 [1.68-11.1]**	3.18 [0.91-11.6]	22.2 [0.44-1917]	3.07 [0.47- 23.0]	2.57 [0.59-12.0]
	Adjusted	1.35 [1.18-1.56]	**2.75 [0.97-7.10]**	1.86 [0.58-7.68]	42.2 [0.41-5259]	3.44 [0.56- 27.1]	1.89 [0.33-9.35]
Mondriaan	Unadjusted	0.75 [0.48-1.17]	0.28 [0.07-1.33]	0.24 [0.02-1.84]	0.09 [0.001- 18.3]	**0.20 [0.01- 3.19]**	**0.44 [0.04-5.43]**
	Adjusted	1.36 [0.84-2.15]	1.19 [0.16- 8.21]	1.10 [0.09-15.4]	0.20 [0.001-93.5]	**0.55 [0.03-16.3]**	**1.67 [0.15-27.7]**

Exposure status is based on the first prescription of a patient. Confounders in the adjusted models are listed in the Table 1. **Bold** numbers indicate that the IV analysis appears valid. SSRI- Selective serotonin re-uptake inhibitor; TCA- Tricyclic antidepressant; THIN- The Health Improvement Network, UK; BIFAP- the Spanish Base de datos para la Investigación Farmacoepidemiológica en Atención Primaria; Mondriaan- Dutch general practices databases; IV- instrumental variable; PSP- proportion of SSRI prescriptions per practice; PSP_{dich}- proportion of SSRI prescriptions per practice (dichotomized at the median); PPP1- single previous prescription by a physician; PPP5- previous five prescriptions by a physician; PPP10- previous ten prescriptions by a physician; CI- confidence interval; *Confidence intervals are estimated by bootstrap method.

those observed for the IV PPP10 (data not shown). For the pooled data, only the IVs PPP5 and PPP10 were strongly associated with the $SSRI_{fixed}$ (r = 0.18 and 0.21, respectively) and confounders were balanced across the categories of these IVs (SDIF < 0.10).

For time-varying exposure, both IVs (PTS and PTS_{dich}) were weakly associated with $SSRI_{time}$ in all databases (Table A1, Appendix). In Mondriaan and BIFAP (PTS_{dich}) confounders were also imbalanced with these IVs (Table A2b, Appendix).

Table 3 shows the hazard ratios (HR) of the association between SSRI use and the risk of HF. In this setting, we used the first prescription ($SSRI_{fixed}$) of eligible patients. In all databases, conventional analysis showed that SSRI use was associated with an increased risk of HF compared to TCA use (though not statistically significant in Mondriaan); the adjusted HR was about 1.35 (Table 3). The results from IV analysis based on the apparently valid IVs (i.e., those IVs that appeared to satisfy the IV assumptions) showed that SSRI use was not associated with an increased risk of HF in any of the databases; the adjusted HRs ranged from 0.55 to 2.75 (Table 3).

The pooled analysis resulted for the conventional analysis in an unadjusted HR of 0.83 [0.78-0.88] and adjusted HR of 1.41[1.33-1.49] and the IV analysis with PPP5 and PPP10 produced unadjusted HRs of 1.10 [0.78-1.53] and 1.07 [0.80- 1.40]; adjusted HRs of 1.70 [1.16-2.50] and 1.57 [1.13-2.15], respectively.

Table 4 shows the HRs for the time-varying exposure ($SSRI_{time}$, current SSRI users vs. non-SSRI users). The conventional analysis showed that the current SSRI use was associated with the increase risk of HF compared to non-SSRI use in all databases; the adjusted HRs were 1.67 [1.56-1.80] (THIN), 1.50 [1.35-1.66] (BIFAP), and 2.49 [1.59-3.90] (Mondriaan) (Table 4). However, because of violation of the key IV assumptions (Table 2), all estimates in this setting showed very different effect estimates (Table 4) compared to the conventional analysis and conclusion from these analyses could not be made.

DISCUSSION

Conventional analysis showed an increased risk of HF for SSRI users versus TCA whereas IV analyses based on the apparently valid IVs (i.e., those IVs that appeared to satisfy the IV assumptions) showed that SSRI users did not indicate a clear association with an increased risk of HF compared to TCA users. However, IV analysis based on the pooled data showed opposite results compared to the analysis per database, i.e., increased risk.

The exposure effects from IV analysis are less precise than those of conventional analysis. None of the IVs we considered appeared to be valid for time-varying exposure ($SSRI_{time}$), and thus the exposure effects are invalid. There is a trade-off between the strength of the IV and the balance of confounders across IV strata (e.g., the IV PPP1 in all databases), which has been described by others.[45] We noticed that results from the conventional analysis seem

Table 4. Associations between time-varying current SSRI versus non-SSRI use and risk of hip fracture based on conventional and IV analysis.

		Conventional§	Instrumental Variables	
			PTS	PTS_{dich}
	Model	HR [CI]	HR [CI]*	HR [CI] *
THIN	Unadjusted	1.76 [1.64-1.89]	0.97 [0.43-2.16]	0.50 [0.20-2.78]
	Adjusted	1.67 [1.56-1.80]	1.90 [0.69-5.19]	0.97 [0.39-4.73]
BIFAP	Unadjusted	2.09 [1.89-2.31]	79.0 [17.5-270]	45.8 [27.1-2083]
	Adjusted	1.50 [1.35-1.66]	4.04 [0.74-16.4]	4.21 [0.54-47.0]
Mondriaan	Unadjusted	1.75 [1.12 -2.72]	0.44 [0.01-17.4]	0.03 [0.001-40.3]
	Adjusted	2.49 [1.59-3.90]	0.83 [0.02-25.8]	0.27 [0.001-32.4]

Exposure status is based on (time-varying) prescription data. Confounders in the adjusted models are listed in the Table 1. SSRI- Selective serotonin re-uptake inhibitor; TCA- Tricyclic antidepressant; Non-SSRI- non-SSRI users could be current, recent or past TCA users or recent/past SSRI users; THIN- The Health Improvement Network, UK; BIFAP- the Spanish Base de datos para la Investigación Farmacoepidemiológica en Atención Primaria; Mondriaan- Dutch general practices databases; IV- instrumental variable; §Time dependent Cox Model. CI- confidence interval; *Confidence intervals are estimated by bootstrap method.

consistent for time fixed and time-varying exposures and across the databases, which is not true for IV analysis mainly due to violations of the key IV assumptions. It is noted that the exposure effects from conventional analysis and IV analysis are not directly comparable as the conventional analysis provided the ACE and the IV analysis provided CACE.[46,47]

With respect to our assessment criteria, we did not identify any IV that is consistently valid across the databases. However, we found that at least one of the IVs for $SSRI_{fixed}$ is valid in each database. For example, the IV PSP in THIN and BIFAP and the IV PPP10 in THIN and Mondriaan are valid because they are strongly associated with the $SSRI_{fixed}$ and confounders are balanced with the categories of these IVs. On the other hand, neither of the IVs across the databases we considered as valid for $SSRI_{time}$. In that case, the main limitation is the weak association between our proposed IVs and the $SSRI_{time}$. We noticed that most of the IVs in THIN and BIFAP are weaker than those of IVs in Mondriaan. This may occur due to little variation in prescribing patterns between practices or prescribing guidelines that influence physician's prescribing behaviour in THIN and BIFAP. These findings indicate that it is still challenging to define a valid IV in pharmacoepidemiological studies especially when exposure and confounders are time-varying. Therefore, future research could be carried out to define an alternative IV in time-varying setting.

We found when the IVs explain a small proportion of the variance of the SSRI exposure (i.e., weakly associated with the exposure, e.g., PPP1 in the BIFAP), the HRs were unstable

with wide confidence intervals.[18] Moreover, these weak IVs may amplify biases due to small violations of the assumptions 2 and 3, and thus IV analysis produces more bias results than the conventional analysis.[11,16,19] We used an arbitrary cut-off (i.e. r < 0.15 and OR < 2) to identify a weak association between IV and exposure. Appropriate cut-off values may differ between studies.[18]

In the analysis of the pooled data, the variation of preferences between practices increased (compared to the analysis per database), resulting in a stronger IV and hence IV estimates are more precise and more stable. Moreover, as IV analysis produces unstable estimates when the outcome is rare,[18] pooling databases may overcome this limitation. Therefore, pooling databases in IV analysis seems effective provided that study protocols are consistent across the different databases.

When the measured confounders are balanced across IV categories (i.e., SDif < 0.10), unmeasured confounders could be balanced as well, in which case assumption 3 is fulfiled.[2,13,27] In that case, we argue that the assumption 2 (IV affects the outcome only through the exposure) might also be fulfilled as the assumption 2 and 3 are statistically similar.[19] Moreover, it is unlikely that a physician's/practice previous antidepressants prescription directly influences their next patient's HF.[20] Additionally, our assumption was that physicians/practices act completely the same (equal standard of care) except for the preference to prescribe either SSRI or TCA. However, if concomitant treatments of depression influence the HF or physician prescribing preference is thought to change during the follow-up, these assumptions may be violated.[2,29,44]

Our study has several strengths. We studied a single drug-event pair (antidepressant drugs-HF) across multiple databases in different countries using a common study protocol to assess the performance of IV analysis with several plausible IVs.[48] Moreover, to our knowledge, this is the first study in which IVs (i.e., proportion of SSRI prescriptions or previous prescriptions by a physician under a practice) are used in time-fixed as well as time-varying exposures and assessed consistently across the databases. Additionally, as we followed a common study protocol, our data extraction was consistent for all three databases; consequently, the design, exposure, outcome, and confounder definitions were harmonized.

There are some limitations in our study. Although we applied a robust balance measure (standardized difference)[39,49] to falsify assumption 3, this could fail to identify a valid IV even when the assumption is violated.[19] For example, we identified the IVs, PSP in BIFAP or PPP10 in Mondriaan, that are valid for the exposure $SSRI_{fixed}$, however, the unadjusted and adjusted IV estimates are very different, which may indicate a possible association between the IV and the unmeasured confounders.[50] Moreover, we provided some theoretical explanations in favour of assumption 2 as it is not possible to verify from the data,[16] however, this assumption may fail even though the assumptions 1 and 3 are fulfilled. Additionally, we have only information of GPs prescriptions; however, we do not know whether patients actually collected their medicines from the pharmacy and took them as prescribed.

We present several recommendations based on our findings: i) as there is a trade-off between the strength of the IV and the balance of the confounders between IV categories, to identify an optimal IV (i.e., strongly associated with exposure under study and independent of the confounders), it would be worthwhile to consider several plausible IVs and assess their validity, 2) if an IV violates one of the assumptions, the IV should be treated as invalid and one should not proceed with IV analysis, on the other hand, if an IV satisfies all assumptions, the IV modelling assumptions should be justified and any inconsistency of the modelling assumptions reported, which helps to understand the validity of the IV analysis.

In conclusion, we assessed the performance of the IV analysis using several potentials IVs in three GP databases using a common study protocol. The performance of IV analysis varied between time-fixed and time-varying exposures, across the databases, and strongly depends on the definition of IVs. Our multiple databases study shows that it is still challenging to obtain a valid IV in pharmacoepidemiological studies especially for time-varying exposure, and thus the exposure effects from IV analysis should be interpreted cautiously.

REFERENCES

1. Groenwold RHH, Hoes A, Nichol KL, Hak E. Quantifying the potential role of unmeasured confounders: The example of influenza vaccination. *Int J Epidemiol* 2008; 37(6): 1422-9.
2. 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; 17(3): 268-75.
3. Ginzburg R, Rosero E. Risk of fractures with selective serotonin-reuptake inhibitors or tricyclic antidepressants. *Ann Pharmacother* 2009; 43(1): 98-103.
4. Thapa PB, Gideon P, Cost TW, Milam AB, Ray WA. Antidepressants and the risk of falls among nursing home residents. *N Engl J Med* 1998; 339(13): 875-82.
5. Van Den Brand MWM, Samson MM, Pouwels S, et al. Use of anti-depressants and the risk of fracture of the hip or femur. *Osteoporosis Int* 2009; 20(10): 1705-13.
6. Eom C-, Lee H-, Ye S, Park SM, Cho K-. Use of selective serotonin reuptake inhibitors and risk of fracture: A systematic review and meta-analysis. *Journal of Bone and Mineral Research* 2012; 27(5): 1186-95.
7. Liu B, Anderson G, Mittmann N, To T, Axcell T, Shear N. Use of selective serotonin-reuptake inhibitors or tricyclic antidepressants and risk of hip fractures in elderly people. *Lancet* 1998; 351(9112): 1303-7.
8. Pouwels S, Van Staa T, Egberts A, Leufkens H, Cooper C, De Vries F. Antipsychotic use and the risk of hip/femur fracture: a population-based case–control study. *Osteoporosis Int* 2009; 20(9): 1499-506.
9. Davies NM, Smith GD, Windmeijer F, Martin RM. Issues in the reporting and conduct of instrumental variable studies: A systematic review. *Epidemiology* 2013; 24(3): 363-9.
10. Rassen JA, Schneeweiss S, Glynn RJ, Mittleman MA, Brookhart MA. Instrumental variable analysis for estimation of treatment effects with dichotomous outcomes. *Am J Epidemiol* 2009; 169(3): 273-84.
11. Ionescu-Ittu R, Abrahamowicz M, Pilote L. Treatment effect estimates varied depending on the definition of the provider prescribing preference-based instrumental variables. *J Clin Epidemiol* 2012; 65(2): 155-62.

12. 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; 45(10): S116-22.
13. Davies NM, Smith GD, Windmeijer F, Martin RM. COX-2 selective nonsteroidal anti-inflammatory drugs and risk of gastrointestinal tract complications and myocardial infarction: An instrumental variable analysis. *Epidemiology* 2013; 24(3): 352-62.
14. Pratt N, Roughead EE, Ryan P, Salter A. Antipsychotics and the risk of death in the elderly: An instrumental variable analysis using two preference based instruments. *Pharmacoepidemiol Drug Saf* 2010; 19(7): 699-707.
15. Huybrechts KF, Gerhard T, Franklin JM, Levin R, Crystal S, Schneeweiss S. Instrumental variable applications using nursing home prescribing preferences in comparative effectiveness research. *Pharmacoepidemiol Drug Saf* 2014; 23(8): 830-8. doi:10.1002/pds.3611.
16. Hernán MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? *Epidemiology* 2006; 17(4): 360-72.
17. Groenwold RHH, Hak E, Klungel OH, Hoes AW. Instrumental Variables in Influenza Vaccination Studies: Mission Impossible?! *Value Health* 2010; 13(1): 132-7.
18. Uddin MJ, Groenwold RHH, de Boer A, Belitser S, Roes KCB, Hoes AW, Klungel OH. Performance of instrumental variable methods in cohort and nested case–control studies: a simulation study. *Pharmacoepidemiol Drug Saf* 2014; 23: 165–177.
19. Swanson SA, Hernán MA. Commentary: How to report instrumental variable analyses (suggestions welcome). *Epidemiology* 2013; 24(3): 370-4.
20. Davies NM, Gunnell D, Thomas KH, Metcalfe C, Windmeijer F, Martin RM. Physicians' prescribing preferences were a potential instrument for patients' actual prescriptions of antidepressants. *J Clin Epidemiol* 2013; 66(12): 1386-96.
21. Suh HS, Hay JW, Johnson KA, Doctor JN. Comparative effectiveness of statin plus fibrate combination therapy and statin monotherapy in patients with type 2 diabetes: use of propensity-score and instrumental variable methods to adjust for treatment-selection bias. *Pharmacoepidemiol Drug Saf* 2012; 21(5): 470-84.
22. Chen H, Mehta S, Aparasu R, Patel A, Ochoa-Perez M. Comparative effectiveness of monotherapy with mood stabilizers versus second generation (atypical) antipsychotics for the treatment of bipolar disorder in children and adolescents. *Pharmacoepidemiol Drug Saf* 2014; 23(3): 299-308.
23. Rassen JA, Brookhart MA, Glynn RJ, Mittleman MA, Schneeweiss S. Instrumental variables II: instrumental variable application-in 25 variations, the physician prescribing preference generally was strong and reduced covariate imbalance. *J Clin Epidemiol* 2009; 62(12): 1233-41.
24. Fang G, Brooks JM, Chrischilles EA. Comparison of instrumental variable analysis using a new instrument with risk adjustment methods to reduce confounding by indication. *Am J Epidemiol* 2012; 175(11): 1142-51.
25. Ahern TP, Pedersen L, Sværke C, Rothman KJ, Sørensen HT, Lash TL. The association between vitamin K antagonist therapy and site-specific cancer incidence estimated by using heart valve replacement as an instrumental variable. *Am J Epidemiol* 2011; 174(12): 1382-90.
26. Kivimäki M, Vahtera J, Kawachi I, et al. Psychosocial work environment as a risk factor for absence with a psychiatric diagnosis: An instrumental-variables analysis. *Am J Epidemiol* 2010; 172(2): 167-72.
27. 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; 54(11): 3390-8.

28. 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. *Can Med Assoc J* 2007; 176(5): 627.
29. Hennessy S, Leonard CE, Palumbo CM, Shi X, Ten Have TR. Instantaneous preference was a stronger instrumental variable than 3- and 6-month prescribing preference for NSAIDs. *J Clin Epidemiol* 2008; 61(12): 1285-8.
30. Julius Clinical Primary Care Research Network. [cited 2014 January, 8]. Available from: http://www.juliusclinical.com/global-presence - cat10.
31. Top Institute Pharma TN. [cited 2014 Januar, 8]. Available from: http://www.tipharma.com/projects/efficiency-analysis-drug-discovery-process/the-mondriaan-project.html.
32. Stirbu-Wagner I, Dorsman S, Visscher S, et al. Landelijk Informatienetwerk Huisartsenzorg. Feiten en cijfers over huisartsenzorg in Nederland [Netherlands Information Network of General Practice. Facts and figures about general practice care the Netherlands]. *Utrecht/Nijmegen: NIVEL/IQ* 2010.
33. ENCePP-. Use of antidepressants and risk of hip/femur fracture. A methodological comparison across data sources and epidemiological design. Electronic Register of Studies, 2014. Available from: http://www.encepp.eu/encepp/openAttachment/fullProtocolLatest/5847;jsessionid=mnbHT9QfbfCJSQLnPl91LQkW3MhQ61rhRXG9gtx23tT8hvCLzcDF!-1931278576 (Date of access: 16 July 2014).
34. Gardarsdottir H, Souverein PC, Egberts TCG, Heerdink ER. Construction of drug treatment episodes from drug-dispensing histories is influenced by the gap length. *J Clin Epidemiol* 2010; 63(4): 422-7.
35. Ali MS, Groenwold RH, Pestman WR, Belitser SV, Hoes AW, de Boer A, Klungel OH. Time-dependent propensity score and collider-stratification bias: an example of beta2-agonist use and the risk of coronary heart disease. *Eur J Epidemiol* 2013; : 1-9.
36. De Vries F, Pouwels S, Bracke M, Lammers J, Klungel O, Leufkens H, Van Staa T. Use of β2 agonists and risk of acute myocardial infarction in patients with hypertension. *Br J Clin Pharmacol* 2008; 65(4): 580-6.
37. Ali M, Uddin MJ, Groenwold RHH, et al. Quantitative Falsification of Instrumental Variables Assumption Using Balance Measures. *Epidemiology* 2014; 25(5): 770-2.
38. Franklin JM, Rassen JA, Ackermann D, Bartels DB, Schneeweiss S. Metrics for covariate balance in cohort studies of causal effects. *Stat Med* 2013; 33(10): 1685-99.
39. Austin PC. A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003. *Stat Med* 2008; 27(12): 2037-49.
40. Angrist JD, Krueger AB. Instrumental variables and the search for identification: From supply and demand to natural experiments. *J Econ Perspect* 2001; 15(4): 69-85.
41. Angrist JD. Estimation of limited dependent variable models with dummy endogenous regressors. *J Bus Econ Stat* 2001; 19(1): 2-28.
42. R Development Core Team. R: A language and environment for statistical computing. *R Foundation for Statistical Computing: Vienna, Austria* 2010; ISBN 3-900051-07-0(01/19).
43. Imbens GW, Angrist JD. Identification and estimation of local average treatment effects. *Econometrica* 1994; : 467-75.
44. Baiocchi M, Cheng J, Small DS. Instrumental variable methods for causal inference. *Stat Med* 2014; 15(33(13)): 2297-340. doi:10.1002/sim.6128.
45. Martens EP, Pestman WR, De Boer A, Belitser SV, Klungel OH. Instrumental variables: Application and limitations. *Epidemiology* 2006; 17(3): 260-7.
46. Fang G, Brooks JM, Chrischilles EA. Apples and oranges? Interpretations of risk adjustment and instrumental variable estimates of intended treatment effects using observational data. *Am J Epidemiol* 2011; 175(1): 60-5. doi:10.1093/aje/kwr283.

47. Boef AG, van Paassen J, Arbous MS, Middelkoop A, Vandenbroucke JP, le Cessie S, Dekkersa OM. Physician's Preference-based Instrumental Variable Analysis- Is It Valid and Useful in a Moderate-sized Study? *Epidemiology* 2014; 10.1097/EDE.0000000000000151:00.
48. ENCePP. European Network of Centres for Pharmacoepidemiology and Pharmacovigilance 2014 [2014/07/16]. Available from: http://www.encepp.eu/encepp/openAttachment/fullProtocolLatest/3690.
49. Belitser SV, Martens EP, Pestman WR, Groenwold RH, de Boer A, Klungel OH. Measuring balance and model selection in propensity score methods. *Pharmacoepidemiol Drug Saf* 2011; 20(11): 1115-29.
50. Brookhart MA, Rassen JA, Schneeweiss S. Instrumental variable methods in comparative safety and effectiveness research. *Pharmacoepidemiol Drug Saf* 2010; 19(6): 537-44.

APPENDIX

Table A1. Instrumental variable assumption 1: association between instrumental variable and SSRI exposure

Instrumental variables	Measures of strength	THIN	BIFAP	Mondriaan
Cohort: first prescription				
Proportion of SSRI prescriptions (continuous, PSP)	Correlation	0.20	0.15	0.26
Proportion of SSRI prescriptions (binary, PSP_{dich})	Odds Ratio	**1.97**	**1.83**	2.44
Previous 1 prescription (PPP1)	Odds Ratio	**1.32**	**1.24**	**1.49**
Previous 5 prescriptions (PPP5)	Correlation	**0.12**	**0.07**	0.18
Previous 10 prescriptions (PPP10)	Correlation	0.15	**0.09**	0.21
Cohort: time-varying settings				
Proportion of time for current SSRI users (continuous, PTS)*	Correlation	**0.09**	**0.08**	**0.14**
Proportion of time for current SSRI users (binary, PTS_{dich})*	Odds Ratio	**1.41**	**1.29**	**1.57**

SSRI- Selective serotonin re-uptake inhibitor; *instrumental variables for time-varying setting, **Bold** numbers indicate the violation of IV assumption 1. THIN- The Health Improvement Network, UK; BIFAP- the Spanish Base de datos para la Investigación Farmacoepidemiológica en Atención Primaria; Mondriaan- Dutch general practices databases.GP- general practice.

Table A2a. Instrumental variable assumption 3: independence between instrumental variable and confounders

	THIN*				BIFAP*				Mondriaan*			
	Exposure	Instrumental variables			Exposure	Instrumental variables			Exposure	Instrumental variables		
Confounders	SSRI vs. TCA	PPP1	PSP[1]	PSP_{dich}	SSRI vs. TCA	PPP1	PSP[1]	PSP_{dich}	SSRI vs. TCA	PPP1	PSP[1]	PSP_{dich}
Age	**0.570**	0.027	0.006	**0.113**	**0.125**	0.001	0.041	0.025	**0.443**	0.015	**0.262**	**0.197**
Sex (Female)	0.033	0.002	0.015	0.014	0.040	0.002	0.001	0.007	0.016	0.004	0.014	0.013
Stroke	0.060	0.002	0.015	0.016	0.013	0.006	0.035	0.030	**0.038**	0.016	0.000	0.045
Ischaemic heart disease	**0.135**	0.008	0.013	0.025	0.010	0.000	0.003	0.010	0.089	0.008	0.022	0.077
Dementia	0.072	0.003	0.002	0.001	0.073	0.001	0.008	0.006	-	-	-	
History of fractures	0.011	0.003	0.014	0.028	0.001	0.005	0.019	0.008	0.050	0.017	**0.189**	0.031
Mental disorders	0.079	0.002	0.007	0.006	0.055	0.003	0.014	0.007	**0.112**	0.028	0.030	0.012
Osteoporosis	**0.116**	0.000	0.014	0.010	0.057	0.009	0.000	0.026	**0.102**	0.003	0.051	0.071
Anaemia	0.049	0.006	0.011	0.023	0.017	0.006	0.021	0.002	0.028	0.002	0.035	0.058
DMARD	0.082	0.004	0.004	0.006	0.03	0.001	0.008	0.000	0.092	0.015	0.046	0.032
Benzodiazepines	0.094	0.009	0.048	0.077	**0.263**	0.005	0.003	0.038	0.164	0.023	**0.134**	0.056
Antiepileptic drugs	**0.114**	0.001	0.001	0.012	**0.157**	0.011	0.001	0.011	**0.224**	0.012	0.009	0.014
Antidiabetic drugs	**0.130**	0.001	0.007	0.011	0.025	0.012	0.012	0.001	**0.178**	0.015	0.000	0.082
Glucocorticoids	**0.120**	0.002	0.007	0.025	0.052	0.008	0.024	0.003	**0.172**	0.016	0.009	0.075
Antihypertensive	**0.244**	0.008	0.007	0.032	0.038	0.002	0.006	0.015	**0.213**	0.014	0.055	**0.126**
ACE inhibitors	-	-	-	-	0.021	0.000	0.017	0.011	**0.117**	0.043	0.051	0.105
Angiotensin II antagonists	-	-	-	-	0.007	0.000	0.007	0.042	**0.144**	0.019	0.074	0.051
Diuretics	**0.215**	0.007	0.01	0.033	0.023	0.004	0.014	0.014	**0.192**	0.017	0.023	0.129

Table A2a. Instrumental variable assumption 3: independence between instrumental variable and confounders (*Continued*)

	THIN*				BIFAP*				Mondriaan*			
	Exposure	Instrumental variables			Exposure	Instrumental variables			Exposure	Instrumental variables		
Confounders	SSRI vs. TCA	PPP1	PSP[1]	PSP_{dich}	SSRI vs. TCA	PPP1	PSP[1]	PSP_{dich}	SSRI vs. TCA	PPP1	PSP[1]	PSP_{dich}
Opioids (including Morphine)	**0.400**	0.000	0.015	0.009	**0.268**	0.011	0.015	0.037	**0.426**	0.002	0.007	0.045
Hormone replacement therapy	0.071	0.005	0.003	0.018	0.034	0.006	0.029	0.003	0.029	0.007	0.007	0.015
Other Antidepressant	0.065	0.002	0.004	0.026	0.025	0.009	0.011	0.019	0.096	0.018	0.041	0.026
Antipsychotics	0.031	0.001	0.009	0.004	0.036	0.005	0.017	0.016	0.089	0.012	0.050	0.011
Vitamin-D	0.025	0.001	0.002	0.004	0.069	0.006	0.004	0.011	0.089	0.009	0.033	0.041
Antiparkinson	0.009	0.005	0.005	0.001	0.013	0.006	0.011	0.016	0.035	0.024	0.022	0.017
Statins	**0.207**	0.006	0.037	0.006	0.029	0.004	0.030	0.033	**0.170**	0.010	0.092	**0.111**
Mahalanobis Balance	0.478	0.001	0.011	0.025	0.227	0.001	0.024	0.009	0.468	0.006	0.171	0.063

[1]Two extreme quintile; *Data for the first prescription; SSRI- Selective serotonin re-uptake inhibitor; TCA- Tricyclic antidepressant; THIN- The Health Improvement Network, UK; BIFAP- the Spanish Base de datos para la Investigación Farmacoepidemiológica en Atención Primaria; Mondriaan- Dutch general practices databases; IV- instrumental variable; PSP- proportion of SSRI prescriptions per practice; PSP_{dich}- proportion of SSRI prescriptions per practice (dichotomized at the median); PPP1- single previous prescription by a physician; Confounders are also balanced with IVs using previous 5 and 10 prescriptions; **Bold** numbers indicate that confounders are not balanced between IV categories (IV assumption 3 is violated).

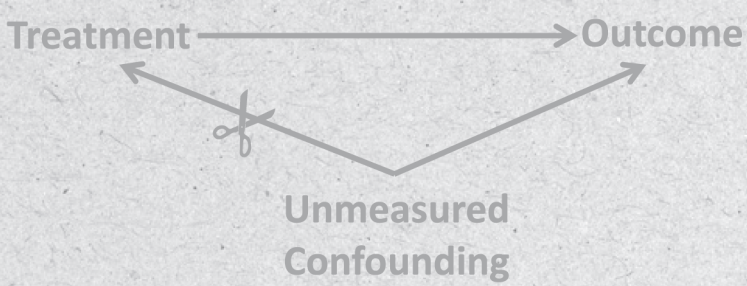
Table A2b. Instrumental variable assumption 3: independence between instrumental variable and confounders

	THIN*		BIFAP*		Mondriaan*	
	Instrumental variables		Instrumental variables		Instrumental variables	
Confounders	PTS[1]	PTS_{dich}	PTS[1]	PTS_{dich}	PTS[1]	PTS_{dich}
Age	0.058	0.047	0.044	**0.145**	**0.106**	**0.257**
Sex (Female)	0.006	0.022	0.008	0.006	0.011	0.010
Stroke	0.009	0.005	0.017	0.005	0.036	0.021
Ischaemic heart disease	0.011	0.017	0.003	0.010	0.084	0.029
Dementia	0.017	0.008	0.013	0.017	-	-
History of fractures	0.028	0.017	0.024	0.019	0.036	**0.134**
Mental disorders	0.005	0.007	0.009	0.005	0.023	**0.104**
Osteoporosis	0.014	0.001	0.004	0.040	0.054	0.091
Anaemia	0.011	0.025	0.009	0.009	0.045	0.020
DMARD	0.005	0.005	0.004	0.007	0.005	0.024
Benzodiazepines	0.046	0.079	0.018	0.075	0.033	0.054
Antiepileptic drugs	0.017	0.012	0.016	0.018	0.024	0.025
Antidiabetic drugs	0.007	0.017	0.016	0.000	0.069	0.001
Glucocorticoids	0.013	0.022	0.010	0.001	0.024	0.015
Antihypertensive	0.026	0.011	0.025	0.042	0.083	0.055
ACE inhibitors	-	-	0.036	0.016	0.048	0.056
Angiotensin II antagonists	-	-	0.002	0.022	0.042	0.023
Diuretics	0.015	0.009	0.081	0.015	0.064	0.062
Opioids (including Morphine)	0.004	0.011	0.003	0.007	0.009	0.031
Hormone replacement therapy	0.009	0.003	0.024	0.01	0.033	0.029
Other Antidepressant	0.015	0.039	0.020	0.039	0.012	0.064
Antipsychotics	0.002	0.012	0.037	0.027	0.015	0.047
Vitamin-D	0.01	0.001	0.010	0.019	0.016	0.021
Antiparkinson	0.008	0.005	0.008	0.026	0.006	0.029
Statins	0.014	0.005	0.036	0.020	0.037	0.052
Mahalanobis Balance	0.019	0.014	0.018	0.031	0.128	0.027

[1]Two extreme quintile; *Data for the time-varying setting; SSRI- Selective serotonin re-uptake inhibitor; TCA- Tricyclic antidepressant; THIN- The Health Improvement Network, UK; BIFAP- the Spanish Base de datos para la Investigación Farmacoepidemiológica en Atención Primaria; Mondriaan- Dutch general practices databases; IV- instrumental variable; PTS- proportion of time for current SSRI users; PTS_{dich}- proportion of time for current SSRI users (dichotomized at the median); **Bold** numbers indicate that confounders are not balanced between IV categories (IV assumption 3 is violated).

Unmeasured confounding IV PERR SCCS Observational studies
PROTECT Causal Inference Statistical methods পরিসংখ্যানগত পদ্ধতি
Statistische methoden IV PERR SCCS Pseudo Randomization
Pharmacoepidemiology IV Analyses PERR SCCS
General Practice Databases Residual Confounding
Electronic Healthcare Databases Epidemiology
Time-fixed Confounding Time-fixed Exposure
Time-varying Exposure
Time-varying Confounding
Non-randomized Studies
IV PERR SCCS
PROTECT 2SLS
IV IV

CHAPTER V
OTHER METHODS TO HANDLE UNMEASURED CONFOUNDING



CHAPTER 5.1

Performance of the Prior Event Rate Ratio Adjustment Method in Pharmacoepidemiology: A Simulation Study

Md Jamal Uddin, Rolf HH Groenwold, Tjeerd P van Staa, Anthonius de Boer, Svetlana V Belitser, Arno W Hoes, Kit CB Roes, Olaf H Klungel

Pharmacoepidemiology and Drug Safety, 2014; DOI: 10.1002/pds.3724

ABSTRACT

BACKGROUND: Prior event rate ratio (PERR) adjustment method has been proposed to control for unmeasured confounding. We aimed to assess the performance of the PERR method in realistic pharmacoepidemiological settings.

METHODS: Simulation studies were performed with varying effects of prior events on the probability of subsequent exposure and post events, incidence rates, effects of confounders, and rate of mortality/dropout. Exposure effects were estimated using conventional rate ratio (RR) and PERR adjustment method (i.e., ratio of RR post exposure initiation and RR prior to initiation of exposure).

RESULTS: In the presence of unmeasured confounding, both conventional and the PERR method may yield biased estimates, but PERR estimates appear generally less biased estimates than the conventional method. However, when prior events strongly influence the probability of subsequent exposure, the exposure effect from the PERR method was more biased than the conventional method. For instance, when the effect of prior events on the exposure was RR = 1.60, the effect estimate from the PERR method was RR = 1.13 and from the conventional method was RR = 2.48 (true exposure effect, RR = 2). In all settings, the variation of the estimates was larger for the PERR method than for the conventional method.

CONCLUSIONS: The PERR adjustment method can be applied to reduce bias due to unmeasured confounding. However, only in particular situations it can completely remove the bias due to unmeasured confounding. When applying this method, theoretical justification using available clinical knowledge for assumptions of the PERR method should be provided.

INTRODUCTION

Unmeasured confounding may impair the validity of observational (pharmaco-) epidemiological studies. A recently proposed method, the prior event rate ratio (PERR) adjustment method, may reduce bias due to unmeasured confounding.[1–5] PERR adjustment is a type of self-controlled design in which the exposure effect is estimated by the ratio of two rate ratios (RRs): RR after initiation of exposure ($RR._{post}$) and the RR prior to initiation of exposure ($RR._{prior}$).[2,6]

A previous simulation study by Yu et al.[1] evaluated and addressed some critical methodological issues of the PERR method. Yu et al. mainly focused on different degrees of association between the confounders and the exposure and different effects of the confounders on the outcome in the prior and post periods. They showed that the PERR method can reduce bias due to unmeasured confounding when the effect of exposure on the outcome is relatively large in comparison with the interaction effects of exposure

and confounders on the outcome or when the time interval effect is rather modest. In addition, this method can reduce bias when unmeasured confounders effects do not vary temporally.[1,2,5]

However, apart from the above stated situations, the PERR adjustment method has not been studied extensively and performance of the PERR adjustment method is unclear in several situations, including situations in which there is an influence of prior events on the probability of subsequent exposure and post events; different incidences of the outcome; different effects of unmeasured confounders, prior events, and exposure on the mortality/dropout in the post period; and different rates of mortality/dropout. Therefore, further methodological investigations into the PERR adjustment method are required to clarify constraints and understand its proper applicability. Our objective was to assess the performance of the PERR method using simulations under various scenarios and to provide guidelines for application of the method.

METHODS

We used Monte Carlo simulations to assess the performance of the PERR adjustment method. First, we briefly describe the PERR adjustment method. Second, the simulation setup, data simulation, and data analysis are described.

Prior event rate ratio (PERR) adjustment method

The PERR adjustment method can be used in a setting where neither the exposed nor unexposed patients are treated with the study drugs before start of the follow-up (Figure 1).[2] In this setting, the rate ratio (RR) observed in the prior period, before initiation of the exposure, is due to differences in patient characteristics between the two study groups. The RR in the post period, after initiation of the exposure, is due to those differences in patient characteristics and exposure. This method requires assumptions about constant temporal effects, i.e., confounding effects are constant across prior and post exposure initiation periods, there is no confounder-by-treatment interaction, and outcomes are nonterminal events.[2,5,7] The exposure effects from the PERR adjustment method is defined as follows:

$$\text{PERR} = \frac{\text{Rate ratio during post period}}{\text{Rate ratio during prior period}} = \frac{\text{RR}_{\text{post}}}{\text{RR}_{\text{prior}}} \quad [1]$$

Throughout the manuscript we will use the abbreviation "PERR" to indicate the exposure effect from the prior event rate ratio adjustment method.

PERR can be estimated either on incidence rate ratios or hazard ratios.[3] Here, the PERR was estimated using incidence rate ratios. Confidence intervals can be obtained by bootstrapping as it is difficult to estimate the covariance between the RRs of the prior and post periods.[1,2,4]

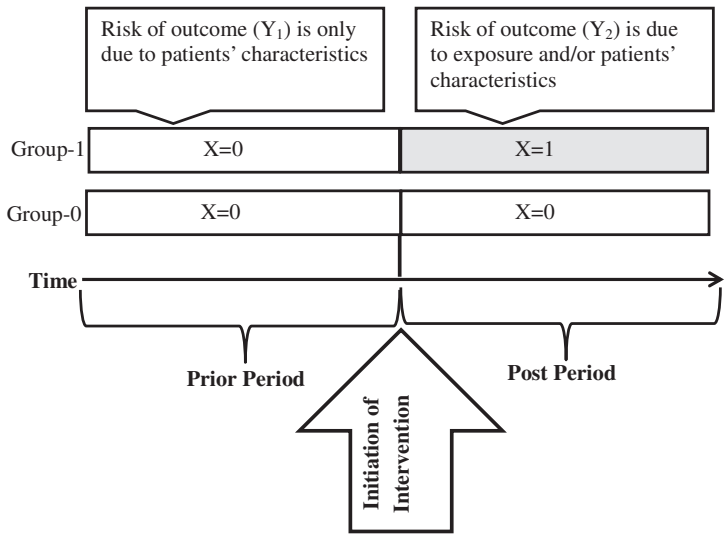


Figure 1. Prior event rate ratio adjustment method.

X = 1 and X = 0 represent the exposed and unexposed subjects, respectively. Y_1 and Y_2 represent the outcomes in the prior and post periods, respectively. The prior period means before intervention and post period means after intervention.

Simulation setup

The following notation is used: X denotes the binary exposure/treatment, subscript number 1 and 2 indicate prior and post periods, respectively, Y_1 and Y_2 denote the outcomes for prior and post periods, C_{11} and C_{12} denote one binary confounder of a subject in prior and post periods, respectively (for example, C_{11} and C_{12} are a measurement of blood pressure of a patient in the prior and post periods, respectively), M_2 denotes the mortality/dropout in the post period. Although we considered binary confounders, performance of PERR is expected to be similar in case of continuous confounders.[1]

We considered several scenarios for the simulation study, which are graphically depicted in causal diagrams. The following scenarios can be identified. Scenario 1: There is confounding in both periods and confounders (C_{11} and C_{12}) are mutually associated as they are from the same patient in the prior and post periods. We assessed the PERR when the confounder effects are constant across both periods. We also evaluated robustness of the method to violation of the assumption of constant effects of confounders in both periods. In that case, the confounder effects are varied between prior and post periods, for example, no differences (effects of C_{11} and C_{12} are constant), small-to-moderate differences (e.g., effect of C_{11} on X and Y_1 is RR 1.5 and effect of C_{12} on X and Y_2 is RR 2.0), and profound differences (e.g., effect of C_{11} on X and Y_1 is RR 2.0 and effect of C_{12} on X and Y_2 is RR 3.5). In this setting, the prior events do not influence the exposure and post events (Figure 2a). In all other

scenarios (i.e., scenarios 2-4), the confounders are associated and effects of confounders are constant across the periods. Scenario 2: Prior events directly influence the post events but do not influence the exposure (Figure 3a). Scenario 3: Prior events influence the probability of subsequent exposure in the absence and presence of confounders (Figure 4a and Figure 4c, respectively). Furthermore, the prior events do not directly influence the post events. Scenario 4: There is mortality/dropout in the post period that is influenced by the confounder (C_{12}), exposure (X), and the prior event (Y_1). In addition, the prior events do not influence the exposure and the post events (Figure 5a).

We also examined the impact of different incidences of the outcomes on the PERR and interaction effects of exposure and confounders (C_{12}) on the outcomes.

In each simulation, the sample size was 100 000 and each scenario was replicated 10 000 times. The incidence rate of the outcome was varied between 1% and 10% in the prior period and 1% and 20% in the post period. The prevalence of exposure was varied between 20% and 25%, and the mortality/dropout rate was varied between 1% and 20%.

Data simulation

Data were simulated as follows. The prior period confounder (C_{11}) follows a binomial distribution with mean probability between 0.20 and 0.30. As the PERR method estimates the 'marginal' effect of exposure on the outcome, we used log-linear models in order to generate our data, to prevent issues of non-collapsibility, because the log-linear model is collapsible.[8,9] In addition, we assumed that the time period is fixed and incidence rates are constant across periods, and hence the estimated rate ratio is equivalent to the risk ratio, which is also collapsible.[10,11]

The log-linear models were used for generating prior events (equation 2), confounder (C_{12}) in the post period (equation 3), exposure (equation 4), post events (equation 5), and mortality or dropout (equation 6). The models are described below:

$$P_{y_1} = exp(\alpha_{y_1} + \beta_{c_1.y_1} C_{11}) \text{ and } Y_1 \sim Bernoulli\ (P_{y_1}) \quad [2]$$

$$P_{c_{12}} = exp(\alpha_{c_{12}} + \beta_{c_{11}.c_{12}} C_{11} + \beta_{y_1.c_{12}} Y_1) \text{ and } C_{12} \sim Bernoulli\ (P_{c_{12}}) \quad [3]$$

$$P_x = exp(\alpha_x + \beta_{c_{11}.x} C_{11} + \beta_{c_{12}.x} C_{12} + \beta_{y_1.x} Y_1) \text{ and } X \sim Bernoulli\ (P_x) \quad [4]$$

$$P_{y_2} = exp(\alpha_{y_2} + \beta_{c_{12}.y_2} C_{12} + \beta_{x.y_2} X + \beta_{y_1.y_2} Y_1) \text{ and } Y_2 \sim Bernoulli\ (P_{y_2}) \quad [5]$$

$$P_{m_2} = exp(\alpha_{m_2} + \beta_{c_{12}.m_2} C_{12} + \beta_{x.m_2} X + \beta_{y_1.m_2} Y_1) \text{ and } M_2 \sim Bernoulli\ (P_{m_2}) \quad [6]$$

where P_{y_1} and P_{y_2} are the probability of the outcome in the prior and post periods, respectively. P_x, P_{m_2}, and $P_{c_{12}}$ are the probability of the exposure, mortality, and confounders (C_{12}), respectively. $\beta_{c_{11}.x}$, $\beta_{c_{11}.y_1}$ and $\beta_{c_{11}.c_{12}}$ are the effects of prior period confounders (C_{11}) on the exposure, prior events, and the post period confounders, respectively and values of these parameters were varied between log(1) and log(10). $\beta_{y_1.c_{12}}$, $\beta_{y_1.x}$, $\beta_{y_1.y_2}$ and $\beta_{y_1.m_2}$ are the effects of prior events on the post period confounders (C_{12}), exposure, post events, and mortality/dropout, respectively and values of these parameters were varied between log(1) and log(5). $\beta_{c_{12}.x}$, $\beta_{c_{12}.y_2}$ and $\beta_{c_{12}.m_2}$ are the effects of post period confounders (C_{12}) on the exposure, post events, and mortality/dropout, respectively and values of these parameters were varied between log(1) and log (5). $\beta_{x.y_2}$ and $\beta_{x.m_2}$ are the effects of exposure on the post events and mortality/dropout, respectively and the values of these parameters were varied between log(1) and log(2). α_{y_1} and α_{y_2} denote the intercepts of the prior and post outcome models, respectively, which indicate baseline incidence rates. α_x denotes the probability (20%) of the exposure among those without prior events and confounder values of zero. $\alpha_{c_{12}}$ and α_{m_2} are the intercepts for the $P_{c_{12}}$ and P_{m_2}. To ensure that the probabilities of $P_{c_{12}}$ and P_{m_2} were between 0 and 1, the values of the parameters $\alpha_{c_{12}}$ and α_{m_2} were varied between -5 and 0.

Data analysis

The PERR adjustment method aims to control for unmeasured confounding. Therefore, in all analyses the confounders (C_{11} and C_{12}) were considered unmeasured. The PERR adjustment exposure effect was estimated using Equation [7a, in the case of mortality/dropout] and Equation [7b, without mortality/dropout] below. The conventional RR was estimated using the data from the post period only, again omitting the (unmeasured) confounder C_{12} from the model. We also analysed data using log-linear model.

The estimation formula of the PERR is as follows:

$$\text{PERR} = \frac{\text{Rate ratio during post period}}{\text{Rate ratio during prior period}} = \frac{E(Y_2 \mid X = 1 \,\&\, M_2 = 0) / E(Y_2 \mid X = 0 \,\&\, M_2 = 0)}{E(Y_1 \mid X = 1) / E(Y_1 \mid X = 0)} \quad [7a]$$

where $M_2 = 0$ denotes whether a subject remained alive in the post period. When there is no mortality/dropout, the formula (7a) can be written as:

$$\text{PERR} = \frac{E(Y_2 \mid X = 1) / E(Y_2 \mid X = 0)}{E(Y_1 \mid X = 1) / E(Y_1 \mid X = 0)} \quad [7b]$$

For the scenario in which the prior events influence the probability of subsequent exposure (i.e., scenario 3), we also estimated the PERR using an adjustment of the observed prior events by inverse probability weighting using the propensity score.

The exposure effects from the PERR and the conventional method were estimated in each simulated dataset, log-transformed, and averaged across datasets. The variation of the estimates was estimated by the standard deviation of the 10 000 estimates. We also estimated 95% confidence intervals in a non-parametric way using the 2.5 and 97.5 percentiles of the 10 000 estimates. We used the statistical software package R (Windows, version 2.15.1) to simulate and analyse the data.[12]

RESULTS

Results based on scenario 1 (in which confounders in both periods were associated, either constant or different across prior and post periods and the prior events did not influence the probability of subsequent exposure and the post events) showed that the conventional RR was biased, RR 2.16 to 4.23 (true RR = 2.00), when the association between confounders and exposure, and confounders and outcome was RR 1.50 to 3.5, respectively. In this setting, the association between the confounders in the prior and post periods was RR =10. When the effects of confounders are constant across prior and post periods, the PERR without mortality/dropout was unbiased (Figure 2b). However, when the effects of confounders are not constant, the PERR was biased (RR 2.27 to 2.91), just as the conventional analysis, but the bias was smaller than for conventional analysis (Figure 2c). The bias for conventional RR and PERR was similar when the confounders in both periods were independent (data not shown).

Figure 3b shows the results for scenario 2, in which prior events directly influence the post events but not the subsequent exposure. The PERR was biased and the bias increased with increasing the relation between prior and post events. For example, the PERR was 2.04 to 2.10 (true RR = 2.00) when the effect of prior events on post events was RR 1.50 to 3.50, respectively. Moreover, the conventional RR was biased but the magnitude of bias was much larger than the PERR method (Figure 3b).

Figure 4 shows the results for scenario 3, in which prior events influence the probability of subsequent exposure in the absence and presence of confounders in both periods, respectively. In the first case, the PERR was highly biased and the bias increased with increasing the effects of prior events on the probability of exposure (Figure 4b). For example, the PERR was 1.53 to 1.11 (true RR = 2.00) when the effect of prior events on exposure was RR 1.25 to 1.60, respectively. However, the conventional RR was unbiased. Similarly, in the second case, the PERR was also highly biased and the bias increased as the effects of prior events increased (Figure 4d). In this case, the conventional RR was also biased, RR = 2.49, (true RR = 2.00). When the prior events strongly influence the probability of the exposure (e.g., RR = 1.60), the bias was more pronounced for PERR (PERR = 1.13) than the conventional RR (RR = 2.48) (Figure 4d). Unlike the bias for PERR in Figure 2c where bias was due to unmeasured confounding and the direction of bias was positive (overestimate the effect), in this case the direction of bias for PERR was negative (underestimate the effect). In scenario 3 with a null exposure effect (RR = 1.00), the PERR shifted away from the null. In a separate

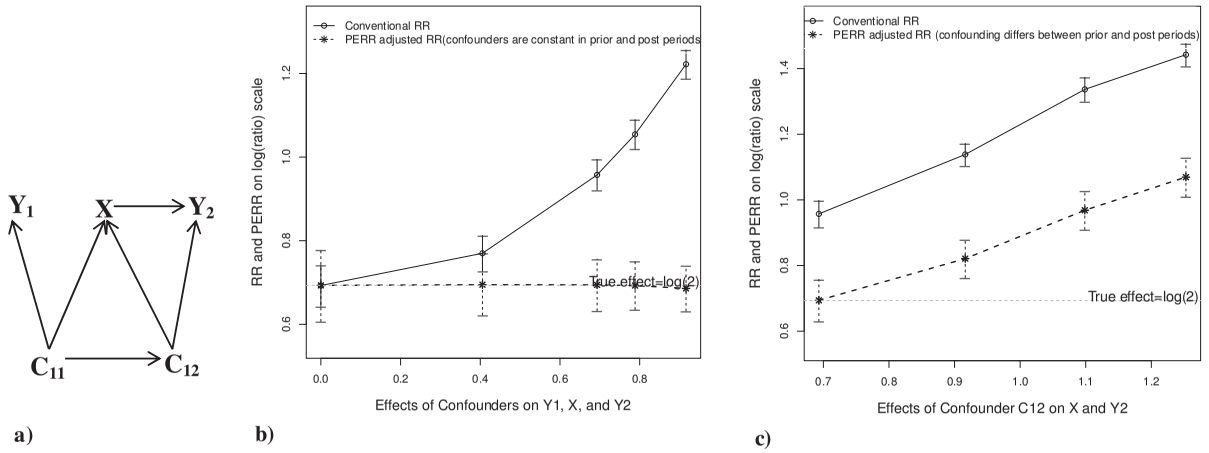


Figure 2. Impact of confounders on prior events rate ratio adjusted estimate and conventional estimate.

2a) Directed acyclic graph that illustrates there are confounders in the both periods (either constant or different) and confounders are associated. Nevertheless, the prior events (Y_1) do not influence the probability of subsequent exposure (X) and the post events (Y_2). X represents the exposure, Y_1 and Y_2 represent the events in the prior and post periods, respectively, and C_{11} and C_{12} represent a confounder in the prior and post periods, respectively. 2b) Effects of confounders on PERR and conventional RR when effects of confounders are constant between prior and post periods. 2c) Effects of confounders on PERR and conventional RR when effect of confounders are different between prior and post periods. RR: rate ratio and PERR: prior event rate ratio. The dotted horizontal straight line (2b-2c) represents the true exposure effect (RR = 2.00, log(2) = 0.6931). Vertical bars indicate the range for 95% of the estimates and are based on the 2.5 and 97.5 percentiles of the distribution of the estimates. Results are based on simulations with sample size 100 000, and each scenario was simulated 10 000 times.

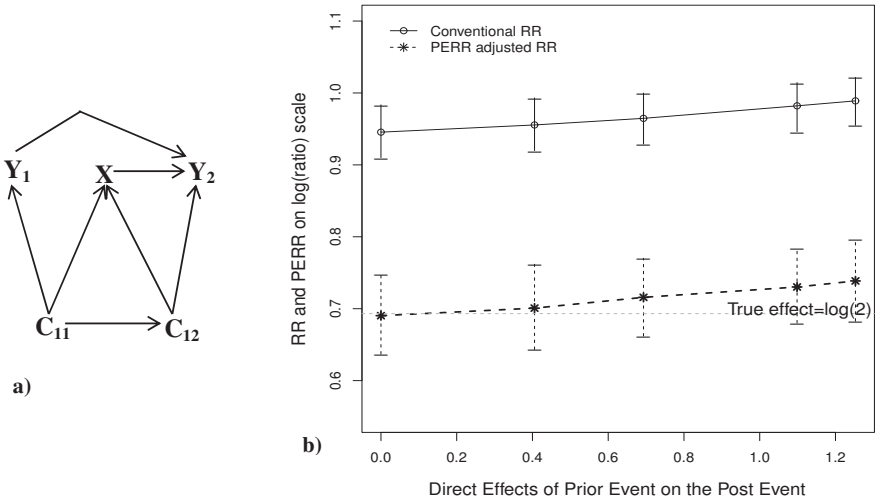


Figure 3. Prior event rate ratio adjusted estimate when prior events (Y_1) directly influence the post events (Y_2).

3a) Directed acyclic graph that illustrates there are confounders in both periods, confounders are associated and prior events directly influence the post events. However, prior event did not influence the exposure. X represents the exposure, Y_1 and Y_2 represent the events in the prior and post periods, respectively, C_{11} and C_{12} represent a confounder in the prior and post periods, respectively. 3b) Effects of prior events (Y_1) on the post events (Y_2). RR: rate ratio and PERR: prior event rate ratio. The dotted horizontal straight line represents the true exposure effect (RR = 2.00, log(2) = 0.6931). Vertical bars indicate the range for 95% of the estimates and are based on the 2.5 and 97.5 percentiles of the distribution of the estimates. Results are based on simulations with sample size 100 000, and each scenario was simulated 10 000 times.

analysis of scenario 3, the PERR was estimated using an adjustment of the observed prior events by inverse probability weighting. This analysis showed that the estimates of the PERR method were similar to the conventional method (Table 1).

Figure 5 shows the results of scenario 4, in which we assessed the impact of an effect of confounders, exposure, and prior events on mortality/dropout. The PERR with taking mortality/dropout into account was more biased (PERR 1.90, rate of mortality = 17% and true RR = 2.00) than the PERR without taking mortality/dropout into account (PERR = 2.00). The bias increased with increasing rates of mortality/dropout.

In all scenarios, the variation of the estimates was more pronounced for PERR than the conventional RR (Table 2). For example, when the effect of prior events on the exposure was RR = 1.50, the standard deviations of the conventional RR and PERR were 0.020 and 0.031, respectively. The variation of the estimates decreased with increasing incidence rates. For instance, when incidence rates were 1% and 20% in the post period, the standard deviations of the PERR were 0.080 and 0.023, respectively.

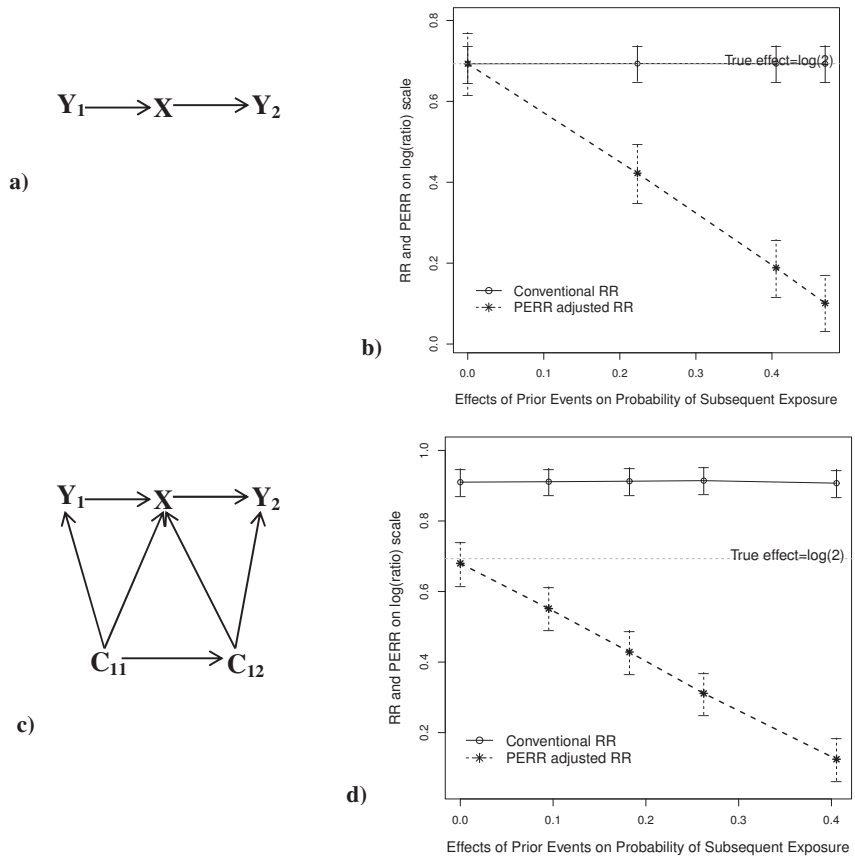


Figure 4. Prior event rate ratio adjusted estimate when prior events (Y_1) influence the probability of subsequent exposure (X).

4a) Directed acyclic graph (DAG) that illustrates there is no confounding variable in both periods but prior events (Y_1) influence the probability of subsequent exposure (X). 4b) Effects of "prior" events on the probability of subsequent exposure in the absence of confounders. 4c) DAG that illustrates there are confounders in both periods; confounders are associated; and the "prior" events influence the probability of subsequent exposure. 4d) Effects of "prior" events on the probability of subsequent exposure in the presence of confounders. X represents the exposure, Y_1 and Y_2 represent the events in the prior and post periods, respectively, and C_{11} and C_{12} represent a confounder in the prior and post periods, respectively. RR: rate ratio and PERR: prior event rate ratio. The dotted horizontal straight line (4b and 4d) represents the true exposure effect (RR = 2.00, log(2) = 0.6931). Vertical bars indicate the range for 95% of the estimates and are based on the 2.5 and 97.5 percentiles of the distribution of the estimates. Results are based on simulations with sample size 100 000, and each scenario was simulated 10 000 times.

We also simulated and analysed data in a situation where interaction effects of exposure and unmeasured confounders were present; we observed that the interaction effects induced bias in the PERR. When prior events did not influence the probability of exposure and post events, and the impact of unmeasured confounding was constant between prior and post

periods, and there was no interaction between exposure and unmeasured confounder, PERR showed unbiased results.

Table 1. Comparison between conventional rate ratio (RR) and both propensity score adjusted PERR (using an adjustment of observed prior event by inverse probability weighting) and unadjusted PERR

Effects of prior event on subsequent exposure in the absence of confounding effect and prior event does not influence the post event (i.e., scenario 3, Figure 4a)	PERR (unadjusted) (in log ratio scale)	PERR (propensity score adjusted) (in log ratio scale)	Conventional RR (in log ratio scale)
Log(1.00)	0.6970	0.6909	0.6902
Log(1.25)	0.4387	0.6935	0.6940
Log(1.50)	0.2046	0.6913	0.6905
Log(1.60)	0.1367	0.6920	0.6920
Effects of prior event on subsequent exposure in the presence of confounding effects but prior event does not influence the post event (i.e., scenario 3, Figure 4c)	**PERR (unadjusted) (in log ratio scale)**	**PERR (propensity score adjusted) (in log ratio scale)**	**Conventional RR (in log ratio scale)**
Log(1.00)	0.6831	0.9121	0.9125
Log(1.10)	0.5512	0.9094	0.9098
Log(1.20)	0.4308	0.9128	0.9145
Log(1.50)	0.2249	0.8830	0.8871
Log(1.60)	0.1399	0.8801	0.8838

Results are based on simulations with sample size 100 000, and each scenario was simulated 10 000 times. The true exposure effect is RR = 2 (log(RR) = 0.6931). PERR: prior event rate ratio.

Table 2. Comparison between standard deviation (SD) of conventional rate ratio (RR) and the PERR adjusted RR

Different effects of prior events on the exposure in the presence of confounders	Conventional RR		PERR adjusted RR	
	Estimate (in log ratio scale)	Standard Deviation *	Estimate (in log ratio scale)	Standard Deviation *
log(RR=1.00)	0.9100	0.0195	0.6798	0.0319
log(RR=1.10)	0.9113	0.0195	0.5515	0.0310
log(RR=1.20)	0.9127	0.0194	0.4284	0.0309
log(RR=1.30)	0.9143	0.0194	0.3112	0.0303
log(RR=1.50)	0.9072	0.0199	0.1246	0.0308

Results are based on simulations with sample size 100 000, and each scenario was simulated 10 000 times.* Standard deviation of the 10 000 estimates. The true exposure effect is RR = 2 (log(2) = 0.6931).

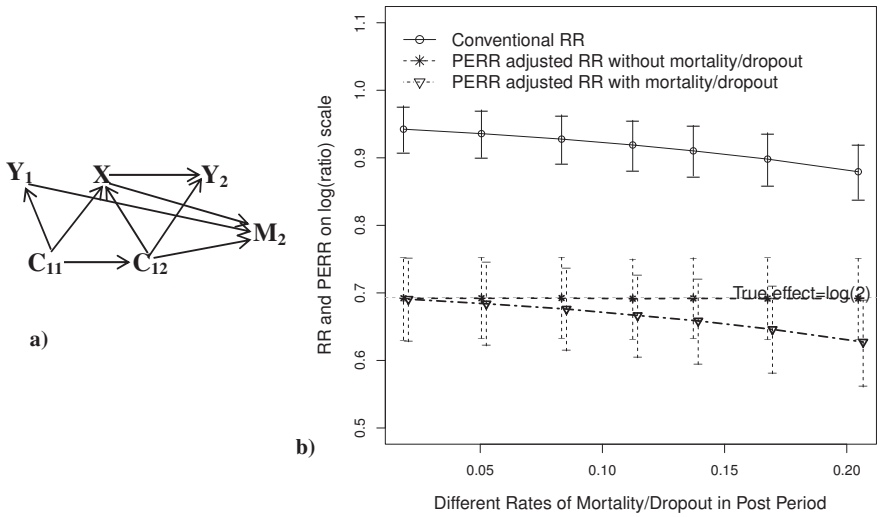


Figure 5. Prior event rate ratio adjusted estimate and conventional estimate when there is mortality or dropout.

5a) Directed acyclic graph that illustrates there is mortality or dropout (M_2) that is influenced by the confounders (C_{12}), prior events (Y_1), and exposure (X); confounders in both periods are associated and constant; the prior event does not influence the probability of subsequent exposure. X represents the exposure, Y_1 and Y_2 represent the outcome in the prior and post periods, respectively, and C_{11} and C_{12} represent a confounder in the prior and post period, respectively, M_2 represents the mortality or dropout in the post period. 5b) Impact of mortality or dropout on the PERR and conventional RR. RR: rate ratio and PERR: prior event rate ratio. The dotted horizontal straight line represents the true exposure effect (RR = 2.00, log(2) = 0.6931). Vertical bars indicate the range for 95% of the estimates and are based on the 2.5 and 97.5 percentiles of the distribution of the estimates. Results are based on simulations with sample size 100 000, and each scenario was simulated 10 000 times. The mortality/dropout rate is about 5-20%.

DISCUSSION

Our simulation study shows that in the presence of unmeasured confounding the PERR adjustment method and conventional regression analysis both yield biased estimates of the treatment effect. However, PERR method results in less biased estimates than the conventional method when confounding differs considerably between prior and post periods, or when prior events directly influence the post events, or when there is mortality or dropout. When prior events influence the probability of the subsequent exposure (e.g., rate ratio ≥ 1.20), the bias is more pronounced for PERR method than the conventional method. Moreover, the bias increased with increasing strength of the relation between prior events and exposure.

In our simulations, the magnitude of bias was smaller for the PERR method than the conventional method when the effects of the unmeasured confounder differed between the prior and post periods (i.e., violation of the assumption of constant confounding effect

in both periods), but again both methods were biased. This is an important finding, because the PERR method has been proposed as a method to deal with unmeasured confounding.[1–5] If the assumption of constant confounding effects is violated, the PERR method may thus yield biased estimates.

The PERR that is estimated by considering mortality/dropout was more biased than those of PERR without taking mortality/dropout into account. The magnitude of bias increases as the rate of mortality/dropout increases. We found the interaction effects of exposure and unmeasured confounder induced bias in the PERR, which is in line with the study of Yu et al.[1] Yu et al. evaluated both PERR and an alternative formulation of the PERR (PERR-ALT), but we only focused on the PERR. The PERR-ALT is a ratio of two RRs: the RR between the post and prior periods for the exposed group and the RR between the post and prior periods for the unexposed group.[1] Because we ignored a setup of clustering in our data and covariates in the analysis, both PERR and PERR-ALT methods provide similar results.

In all settings, the variation of the estimates was larger for PERR than the conventional RR. This may due to the fact that PERR is the ratio of two RRs between the prior and the post periods. The variation decreases as incidence rates increases in both periods. Tannen et al[2] stated that the variation in the PERR can be larger with smaller number of prior events. We argue that if researchers believe that the conventional point estimates are substantially biased due to unmeasured time-fixed confounding, the PERR adjustment method can be a viable alternative if assumptions are met. Furthermore, because the PERR method yields large standard errors compared to conventional methods, the method can be considered more conservative than conventional methods.

We observed that both the magnitude of bias and its direction vary with the effects of prior events on the probability of exposure as well as the different effects of confounders between prior and post periods. For a given effect of prior event and confounders, the bias may be either negative or positive, which is related to the ratio of the two RRs. More clearly, if the RR before initiation of the exposure is larger than the RR after initiation of the exposure, the denominator of the PERR is larger than the numerator, the resulting exposure effect is moving to the null effect and bias becomes negative. In that case, when prior event strongly influences the probability of the subsequent exposure, the exposure effect is very close to the null effect. A similar explanation applies to positive bias when bias is due to the unmeasured confounders.

The major limitation of the PERR method is the requirement that the event rate prior to study start can be estimated. Thus this method cannot be applied when patients have no records prior to the study start, or for patient who die prior to exposure initiation.[1,4] Apart from these limitations, the findings of our simulation study suggest that PERR method is not an adequate method to remove bias due to unmeasured confounding in situations when prior events influence the probability of subsequent exposure even when there is no exposure effects on the outcome. An example of likely bias with PERR would be the study

of statin use and risk of myocardial infarction (MI) or stroke. When a patient first experienced a MI or stroke, (s)he has to start treatment (e.g., statin) in order to avoid a secondary MI or stroke, and consequently the health characteristics of this patient (e.g., cholesterol levels) are influenced by such type of therapy. Hence, the probability of the exposure (start statin therapy) is definitely influenced by previous outcome as well as the confounders.

The PERR method may be a valid method to remove bias due to unmeasured confounding in a situation where the probability of exposure and probability of post events are independent of the prior events. An example may be a study of the effect of proton-pump inhibitors (PPIs) on the risk of pneumonia. Patients may experience a pneumonia multiple times in their life; thus it can occur before and/or after initiation of the PPIs, hence one of the prerequisites to apply the PERR method (outcome before intervention) is met. Moreover, if a patient has suffered pneumonia before initiation of the PPIs, this prior outcome may not influence the likelihood of being prescribed PPIs, because PPIs are generally prescribed to relieve symptoms of gastroesophageal reflux disease and not for pneumonia. Hence, another important condition (independence between prior event and subsequent exposure) of the PERR adjustment method is satisfied here. However, in this setting the PERR method may provide biased results when the unmeasured confounder effects vary temporally in prior and post periods and prior events directly influence the post events. Another example where the PERR method could also be valid is a study of statin use and risk of hip fracture or lower urinary tract symptoms. Like in the PPIs example, this example may also satisfy the critical assumptions underlying the PERR method. However, if the rate of mortality or dropout is high, or prior events influence the post events, or the unmeasured confounder effects vary temporally, the PERR adjustment method may not provide unbiased results.

We note that the PERR adjustment method has some similarities with the case-time-control (CTC) design and case-crossover design (CCD).[13–15] For example, all three methods deal with intermittent exposures with transient effects, and overcome confounding by constant characteristics.[1,13,15,16] However, there are some differences between these methods. Differences between the CCD and the PERR methods are that in the PERR method both cases and non-cases are used, whereas CCD only uses information on cases. Furthermore, in the CCD, exposure misclassification is more likely and bias may arise in selection of the control time window(s)[15] but not in the PERR method. In addition, the PERR method generally deals with time-to-event outcome but CCD deals with binary outcomes and is analysed using conditional logistic regression to account for the matched nature of the data.[17] A possible limitation of the CCD is that it does not account for general time trend in drug use,[13] while this limitation does not appear in the PERR. Differences between the CTC and the PERR methods include that CTC also deals with binary outcome, but the PERR generally deals with time-to-event outcome.[1,13] The CTC assumes common period effect for cases and controls, whereas the PERR does not make this assumption.[1,13]

For the simplest situation in Figure 2, the prior event Y_1 acts just like a negative control (i.e., an outcome such that the set of unmeasured confounders C_{12} of exposure X and post

event Y_2 should be as identical as possible to the set of unmeasured confounders C_{11} of exposure X and prior event Y_1)[18] for detecting unmeasured confounding. If the negative control is empirically associated with the exposure after adjustment for possible measured confounders, unmeasured confounding may be present in the data.[18–21] The prior event also acts as a 'perturbation variable' (a variable that is associated with the exposure and post event only through unmeasured confounders) to detect and correct (via adjustment of the perturbation variable) the bias of unmeasured confounding.[18–20] Detailed explanations of negative controls and perturbation variables can be found in Lipsitch et al.,[18] Flanders et al.,[20] Lee.,[19] and Tchetgen Tchetgen.[21]

For a complex scenario, such as when prior events influence the probability of subsequent the exposure (i.e., scenario 3, Figure 4), one can adjust for the prior events (e.g., via inverse probability weighting using propensity score). Once this adjustment is done, the PERR method yields the same result as the conventional method. Although it is easy to assess such complex scenarios in simulations, this may be very hard in empirical studies. Therefore, before going to apply the PERR adjustment method in empirical data, theoretical motivations on the appropriateness of applying the method should be given based on clinical knowledge.

Although Yu et al.[1] indeed pointed out some crucial situations for the PERR adjustment method, our simulations go beyond those and showed that in a particular cases (e.g., the prior event rate strongly influences the probability of exposure) the PERR method provides more biased estimates than the conventional method. Moreover, we show that the PERR method is sensitive to small violation of one of the key assumptions (i.e., constant effects of confounders in both prior and post periods). In addition, we simulated and analysed data in the case of dropout and mortality, which is important in a follow-up study.

Strengths of our simulation studies are that we generated data in several scenarios that are commonly seen in pharmacoepidemiological research; the conventional RR and PERR are estimated by a simple ratio technique, which is easy to understand and free from different assumptions of statistical modelling techniques; and the sample size and number of replications are large enough to achieve enough power. Limitations of our study are that the time effects and incidence rates are constant across prior and post periods. Future research could explore settings of different time interval effects and different incidence rates in both periods. Although we did not estimate coverage probabilities of the PERR, we expect the coverage probabilities are far away from the nominal level (especially in the reported scenarios) due to large amount of bias for the PERR method.

In conclusion, the PERR adjustment method can be applied to reduce bias due to unmeasured confounding. However, in particular situations, e.g., when prior events strongly influence the probability of subsequent exposure, this method can be more biased than conventional methods. Hence, when applying this method, we suggest to provide theoretical justification using available clinical knowledge for underlying assumptions of the PERR method.

REFERENCES

1. Yu M, Xie D, Wang X, Weiner MG, Tannen RL. Prior event rate ratio adjustment: Numerical studies of a statistical method to address unrecognized confounding in observational studies. *Pharmacoepidemiol Drug Saf* 2012; 21(SUPPL.2): 60-8.
2. Tannen RL, Weiner MG, Xie D. Use of primary care electronic medical record database in drug efficacy research on cardiovascular outcomes: Comparison of database and randomised controlled trial findings. *BMJ (Online)* 2009; 338(7691): 395-9.
3. Tannen R, Xie D, Wang X, Yu M, Weiner MG. A new "Comparative Effectiveness" assessment strategy using the THIN database: Comparison of the cardiac complications of pioglitazone and rosiglitazone. *Pharmacoepidemiol Drug Saf* 2013; 22(1): 86-97.
4. Tannen RL, Weiner MG, Xie D. Replicated studies of two randomized trials of angiotensin-converting enzyme inhibitors: further empiric validation of the 'prior event rate ratio' to adjust for unmeasured confounding by indication. *Pharmacoepidemiol Drug Saf* 2008; 17(7): 671-85.
5. Weiner MG, Xie D, Tannen RL. Replication of the Scandinavian Simvastatin Survival Study using a primary care medical record database prompted exploration of a new method to address unmeasured confounding. *Pharmacoepidemiol Drug Saf* 2008; 17(7): 661-70.
6. Schneeweiss S, Seeger JD, Smith SR. Methods for developing and analyzing clinically rich data for patient-centered outcomes research: An overview. *Pharmacoepidemiol Drug Saf* 2012; 21(SUPPL.2): 1-5.
7. Alemayehu D, Alvir JM, Jones B, Willke RJ. Statistical issues with the analysis of nonrandomized studies in comparative effectiveness research. *Journal of managed care pharmacy : JMCP* 2011; 17(9 Suppl A): S22-26.
8. Agresti A. Categorical data analysis. John Wiley and Sons; 2002.
9. Ali MS, Groenwold RHH, Pestman WR, et al. Propensity score balance measures in pharmacoepidemiology: A simulation study. *Pharmacoepidemiol Drug Saf* 2014; 23(8): 802–811.
10. Rothman KJ. Epidemiology: an introduction. Oxford University Press; 2012.
11. Greenland S. Confounding of incidence density ratio in case-control studies. *Epidemiology* 2013; 24(4): 624-5. doi:10.1097/EDE.0b013e3182953c23 [doi].
12. R Development Core Team. R: A language and environment for statistical computing. *R Foundation for Statistical Computing: Vienna, Austria* 2010; ISBN 3-900051-07-0(01/19).
13. Suissa S. The case-time-control design. *Epidemiology* 1995; 6(3): 248-53.
14. Suissa S, Dell'Aniello S, Martinez C. The multitime case-control design for time-varying exposures. *Epidemiology* 2010; 21(6): 876-83.
15. Schneeweiss S, Stürmer T, Maclure M. Case-crossover and case-time-control designs as alternatives in pharmacoepidemiologic research. *Pharmacoepidemiol Drug Saf* 1997; 6(SUPPL. 3): S51-9.
16. Maclure M, Mittleman aM. Should we use a case-crossover design? *Annu Rev Public Health* 2000; 21(1): 193-221.
17. Delaney JA, Suissa S. The case-crossover study design in pharmacoepidemiology. *Stat Methods Med Res* 2009; 18(1): 53-65. doi:10.1177/0962280208092346 [doi].
18. Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative Controls: A tool for detecting confounding and bias in observational studies. *Epidemiology* 2010; 21(3): 383-8.
19. Lee W-. Detecting and correcting the bias of unmeasured factors using perturbation analysis: A data-mining approach. *BMC Medical Research Methodology* 2014; 14:18(1).
20. Flanders WD, Klein M, Darrow LA, et al. A method for detection of residual confounding in time-series and other observational studies. *Epidemiology* 2011; 22(1): 59-67.
21. Tchetgen Tchetgen E. The control outcome calibration approach for causal inference with unobserved confounding. *Am J Epidemiol* 2014; 179(5): 633-40. doi:10.1093/aje/kwt303 [doi].

CHAPTER 5.2

Impact of Violations of the Assumptions of the Self-controlled Case Series Design: a Cautionary Note

Md Jamal Uddin, Rolf HH Groenwold, M Sanni Ali, Gianmario Candore, Mark CH de Groot, Patrick C Souverein, Yolanda Alvarez, Svetlana V Belitser, Arno W Hoes, Anthonius de Boer, Kit CB Roes, Olaf H Klungel, Helga Gardarsdottir

Submitted, Pharmacoepidemiology and Drug Safety

ABSTRACT

BACKGROUND: The self-controlled case-series (SCCS) design has been used to control for time-fixed (un)measured confounding in healthcare databases. We aimed to evaluate the impact of violation of assumptions of the SCCS design and different definitions of observation periods as well as risk periods in two general practice databases in a study of antidepressants use and risk of hip/femur fracture (HF).

METHODS: Information on adults with a hip/femur fracture (HF) and a prescription for antidepressants at any time during the observation period 2001-2009 was extracted from the UK THIN (6632 cases) and the Dutch Mondriaan (136 cases) databases. The incidence rate ratio (IRR) was estimated using conditional Poisson regression.

RESULTS: The IRRs appeared extremely biased when subjects were censored at their first/last HF or when analysis included only subjects whose follow-up started with initiation of antidepressants (i.e., incident users). For example in THIN, IRRs for >365 days of exposure were 1.26 [1.13-1.42] when complete follow-up was considered and 40.1 [32.2-49.9] when censoring is at the first event. However, results were consistent to including subjects who were exposed at the start of follow-up and different risk period definitions.

CONCLUSIONS: The SCCS design is sensitive to violations of the assumptions and yields apparently biased estimates when observations are censored at the event or when follow-up starts with the first prescription. The performance of this design may differ across studies and across databases. Therefore, in each SCCS study, correct specification of the SCCS design should be carefully assessed and reported.

INTRODUCTION

Electronic healthcare databases have been used extensively for observational pharmacoepidemiological studies to identify (un)intended effects of drugs. In such studies, cohort and case-control designs are frequently used. However, confounding by unmeasured patient characteristics (e.g., frailty, severity of disease) may bias results of such studies. For example, unmeasured confounding may (partly) explain the observed adverse effects of antidepressants on the risk of hip/femur fracture (HF).[1–3] The self-controlled case-series (SCCS) design may control for all time-fixed measured and unmeasured confounding and has been increasingly used in pharmacoepidemiology when investigating a transient exposure and an acute event.[4–13]

The SCCS design uses only information from subjects who experienced the event of interest, similar to the case-crossover design in which subjects act as their own controls. This design was originally developed to evaluate vaccine safety[14] and is extensively described in a tutorial by Whitaker et al.[15] In short, individual follow-up time is divided into exposed and

unexposed periods, which may vary in length of duration. A comparison is made within subjects, between the incidence rates of outcome of interest in these periods.[15] The design also uses the exposure history occurring after the event and occurrence of multiple events per subject is allowed in this design.[16]

The three key assumptions underlying the SCCS design are: i) events are allowed to be recurrent and independent or are unique and uncommon over the study periods; ii) the occurrence of an event must not alter the probability of subsequent exposure; and iii) the occurrence of the event should not censor the observation period.[7,14,15,17–19]

Several studies,[4,5,7,8,12–14,19–25] evaluated the SCCS design in various research areas. For example, Hubbard et al.[4] determined the association between antidepressants use and the risk of HF and they compared the SCCS results with a classical case-control design. Moreover, Douglas et al.,[7,8] and Ramsay et al.[13] evaluated the robustness of results from SCCS studies using sensitivity analyses. Although several studies[15,17,19,26–28] acknowledged that violations of the key SCCS assumptions lead to bias exposure effects, little is known whether the impact of the violations is substantial in empirical studies. In addition, the performance of the SCCS design in situations where the SCCS analysis is restricted to incidence user (i.e., subjects whose follow-up started with the initiation of exposure) or prevalent users and different definitions of risk windows in pharmacoepidemiological studies has not been well studied. Therefore, we aimed to evaluate the impact of violations of assumptions of the SCCS design and different definitions of observation periods (e.g., follow-up started with initiation of exposure) as well as risk periods in two general practice databases (Mondriaan, Netherlands and The Health Improvement Network, THIN, UK) in a study of antidepressants use and risk of hip/femur fracture.

METHODS

Data sources, study population and study period

We used data from Dutch Mondriaan and The Health Improvement Network (THIN, UK). Detailed information of the databases can be found elsewhere.[29,30]

The study population included all subjects in the Mondriaan and THIN who, at any time during the study period (1 January 2001 to 31 December 2009) fulfilled the quality standard criteria for each database, were ≥ 18 years old, and had at least one year of enrolment with a general practice. They had to have at least one antidepressant prescription prescribed and a recorded diagnosis of HF during the study period. Once subjects were eligible, at least 12 months should have elapsed without event for a HF to be considered a new event and 6 months without antidepressant use to be considered incident antidepressant use. Subjects who were excluded because of an antidepressants prescription in the 6 months before start of the study period were included in a separate analysis (see later). For subjects with at least one year enrolment at 1 January 2001, the observation period for the main analysis started on 1 January 2001. Subjects who fulfilled the eligibility criteria after 2001 or turned 18 years during the study period could enter at any time in the study given that the

above mentioned inclusion criteria were fulfilled. The end of the observation period was the moment a patient died, left the database or the practice left the database, or the end of the study period, whichever came first.

Exposure and outcome definitions

The duration of antidepressant use was determined by calculating the length of treatment episodes, which were defined as a series of subsequent prescriptions for antidepressants, independent of switching of type and dose change and was constructed according to the method of Gardarsdottir et al.[31] Detailed information of the treatment episodes can be found in the study protocol.[30]

The total exposure times of subjects were divided into periods of current, recent and past use, switching between these periods according to drug use. More clearly, each individual observation time was divided into risk windows as follows: period 0: no use, from start of the observation period until the first antidepressant prescription; period 1: current use, from 0-30 days after start of antidepressants; period 2: current use, from 31-182 days after start of antidepressants; period 3: current use, from 183-365 days after start of antidepressants; period 4: current use, from >365 days after start of antidepressants; period 5: a period of 60 days after the current use (equivalent to recent use), and period 6: a period starting after period 5 until becoming a user again or the end of the study. In the analysis, the periods 0 and 6 were combined into the reference period (no use and past use).

All subjects with a record/diagnosis of a hip/femur fracture and who received antidepressants during the study period were included in the analysis, regardless of what came first, e.g., the hip fracture or the antidepressant use. When the patient had a second hip/femur fracture, a minimum of 12 months should have elapsed between the first and the second fracture in order to consider the second fracture as a new event

Potential confounders

Age was considered as the only confounder in the main analysis and included in the adjusted model as a categorical variable because it is an important risk factor for the outcome and obviously changes when observation periods are long.[19] Age was categorised by a single year band when age < 30 or ≥ 60 years and five years band when age 30-59 years. Information of a subject was updated when 1) exposure status changes, or 2) the value of age changes (i.e., at a subject's birthday). Although the SCCS design control for all time-fixed confounding, it may not account for time-varying confounders such as use of co-medication. In a separate analysis, we also assessed the impact of adjustment for benzodiazepine use.

Data analysis

Conditional Poisson regression model, which assumes that events at different time points are conditionally independent given exposures, was used to estimate the incidence rate ratio (IRR) of HF with the use of antidepressants and with corresponding 95% confidence interval (CI).[15,32] The IRR was defined as the rate of events during exposed periods divided

by the rate during all other observed periods.[7,14,17,18] Duration of follow-up for a particular observation period was included as an offset variable in the model (i.e., log(time)). Two sets of analysis were performed, without and with age adjustment. All statistical analysis was conducted using SAS 9.2 for THIN data and R 3.1.0 for Mondriaan data. Each database owner performed all analyses locally. A blinding procedure was maintained until all results were available at the coordinator centre (at Utrecht University, the Netherland).

To evaluate the impact of the violation of the assumptions of the SCCS design, several analyses were performed. First, we investigated the possible event-exposure dependence by defining a "pre-exposure" period of 30-days prior to initiating of an antidepressant. Second, we evaluated the impact of right censoring before end of the follow-up. Specifically, subjects were right censored at their first event and (in a separate analysis) at their last event during the study period. Third, an analysis was done involving only incident users: i.e., follow-up started with the first antidepressant prescription (antidepressant use before HF), excluding those subjects whose antidepressant use started after their first HF. This analysis was performed in Mondriaan only as similar trends were expected in THIN. Fourth, we assessed the impact of including subjects who were prescribed antidepressants in the 6 months before the start of follow-up.

We also assessed the impact of different exposure definitions on the results. The follow-up time was divided into different observation periods by exposure state: period 1: from 1 to 90 days after initiating antidepressant use, period 2: a combination of the remainder of current use (from 91 days after initiating antidepressant use until the end of the current period) and the period of recent use, period 3: includes the no use (from start of the observation period until the first antidepressant prescription) and the past use (a period starting after the recent use until becoming a user again or the end of the study). Moreover, we studied the effect of different type of antidepressant use defined as: use of a selective serotonin reuptake inhibitor (SSRI), a tricyclic antidepressant (TCA) and both SSRI and TCA. For this analysis, all periods of current use were pooled into a single period: 'current SSRI', 'current TCA', or 'current both SSRI and TCA'. 'Current both SSRI and TCA' use could include combined use or subsequent use within the same treatment episode. In addition, we conducted an analysis in Mondriaan only to assess the impact of adjustment for the potential time-dependent confounder benzodiazepine use. For the definition of the confounder benzodiazepine use, we refer to the protocol.[30] In all analyses, models were age-adjusted.

RESULTS

In total, 136 (Mondriaan) and 6632 (THIN) patients were included for analysis. Characteristics of these subjects are described in Table 1. Table 2 shows the IRR of HF among antidepressant exposure (main analysis). Age-adjusted analysis showed that the largest risk was observed during the period 1-30 days after start of antidepressants use: IRR 3.22 [95% CI 1.51-6.84] in Mondriaan, and IRR 1.57 [1.39-1.78] in THIN.

Table 1. Demographics and characteristics of the study populations

	Mondriaan				THIN			
	All (136)	SSRI exposure during follow-up (79)	TCA exposure during follow-up (69)	SSRI & TCA exposure during follow-up (2)	All (6632)	SSRI exposure during follow-up (3983)	TCA exposure during follow-up (3213)	SSRI &TCA exposure during follow-up (593)
Males (%)	33 (24.3)	19 (24.5)	16 (23.2)	1 (50.0)	1556 (23.5)	942 (23.7)	709 (22.1)	142 (23.9)
Mean age at first exposure (SD)	72.2 (16.4)	73.5 (16.8)	71.4 (15.6)	68.5 (18.7)	76.2 (14.1)	76.5 (14.8)	75.9 (12.8)	73.7 (14.4)
Duration of observation period (median years)*	7.0	7.0	7.1	6.0	8.0	7.7	8.6	8.3
Duration of the current antidepressants exposure (median days)	219	248	194	-	256	252	124	456
Fractures during follow-up	136	79	69	2	6632	3983	3213	593
Fractures during an episode of antidepressants exposure	48	35	13	-	2210	1303	687	220

*Observation period is from start date until the end of observation period, SD- Standard deviation, SSRI- selective serotonin reuptake inhibitor, TCA- tricyclic antidepressant, THIN- The health improvement network

Table 2. Incidence rate ratios of hip/femur fracture based on a SCCS design among antidepressant exposure categories (main analysis)

	Mondriaan			THIN		
Exposure	Cases	IRR [95% CI]	IRR [95% CI] (Age adjusted)	Cases	IRR [95% CI]	IRR [95% CI] (Age adjusted)
Reference period*	82	1 (Ref)	1 (Ref)	4,429	1 (Ref)	1 (Ref)
Current use (1-30 days)	7	3.83 [1.81-8.12]	3.22 [1.51-6.84]	274	1.87 [1.65-2.12]	1.57 [1.39-1.78]
Current use (31-182 days)	24	3.37 [2.15-5.30]	2.76 [1.69-4.50]	731	1.91 [1.75-2.08]	1.52 [1.39-1.65]
Current use (183-365 days)	8	2.78 [1.27-6.09]	1.94 [0.84-4.47]	380	2.10 [1.87-2.36]	1.47 [1.31-1.66]
Current use (>365 days)	9	2.83 [1.20-6.68]	1.61 [0.57-4.49]	825	2.34 [2.11-2.60]	1.26 [1.13-1.42]
Recent use (post-exposure period)	6	2.70 [1.18-6.16]	2.18 [0.97-4.88]	286	1.61 [1.43-1.82]	1.37 [1.21-1.55]

IRR- incidence rate ratios, CI- confidence interval, AD- antidepressants, SCCS- self-controlled case series, THIN- The health improvement network, *Reference period comprises the no use (from start of the observation period until the first antidepressant prescription) and the past use (a period starting after the recent use until becoming a user again or the end of the study).

Table 3 shows the results of the analysis focusing on possible event-exposure dependence. Compared to reference period, there appeared to be an increased risk during the 30 days prior to the start of exposure: IRR 2.51 [1.00-6.33] for Mondriaan and IRR 1.22 [1.06-1.41] for THIN, suggesting event-exposure dependence. The pattern of IRRs in other categories of the exposure was similar with the main analysis.

Table 3. Incidence rate ratios of hip/femur fracture based on a SCCS design using event-exposure dependence analysis

	Event-exposure dependence analysis			
	Mondriaan		THIN	
Exposure	Cases	IRR [95% CI] (Age adjusted)	Cases	IRR [95% CI] (Age adjusted)
Reference period*	77	1 (Ref)	4,264	1 (Ref)
Pre-exposure period, 30 days prior to initiating AD use	5	2.51 [1.00-6.33]	208	1.22 [1.06-1.41]
Current use (1-30 days)	7	3.47 [1.62-7.42]	274	1.60 [1.41-1.81]
Current use (31-182 days)	24	2.97 [1.80-4.90]	723	1.55 [1.42-1.69]
Current use (183-365 days)	8	2.10 [0.91-4.88]	377	1.50 [1.33-1.69]
Current use (>365 days)	9	1.74 [0.62-4.93]	821	1.28 [1.14-1.44]
Recent use (post-exposure period)	6	2.31 [1.03-5.20]	258	1.44 [1.27-1.64]

IRR- incidence rate ratios, CI- confidence interval, AD- antidepressants, SCCS- self-controlled case series, THIN- The health improvement network, *Reference period comprises the no use (from start of the observation period until the first antidepressant prescription) and the past use (a period starting after the recent use until becoming a user again or the end of the study).

Patients experienced only a single event in Mondriaan. When subjects were censored at the first or the last event in THIN or at the event in Mondriaan, IRR increased considerably compared to the main analysis. For example, when subjects were exposed for more than 365 days, IRRs were 68.2 [5.32-875] for Mondriaan and 8.80 [6.92-11.2] for THIN (censoring at the last event) (Table 4). In that case, the observation time that was removed from the analysis was mainly time during which subjects were exposed (for example, in Mondriaan, 69% of the observation time that was removed was exposed observation time).

When including only follow-up time after the first antidepressant prescription (i.e., removing unexposed observation time prior to the first prescription), the estimated IRRs increased. For example, in Mondriaan the IRR for the first 30 days of antidepressant use was 12.5 [5.39-29.0] (Table 5). In this analysis, 54 of 82 cases in the reference group were removed from the analysis but the number of exposed cases obviously remained the same.

Table 4. Incidence rate ratios of hip/femur fracture based on a SCCS design when subjects censor at their first and last events

	Mondriaan** (right censoring at event)		THIN (right censoring at first event)		THIN (right censoring at last event)	
Exposure	Cases	IRR [95% CI] (Age adjusted)	Cases	IRR [95% CI] (Age adjusted)	Cases	IRR [95% CI] (Age adjusted)
Reference period*	82	1 (Ref)	2,044	1 (Ref)	2,259	1 (Ref)
Current use (1-30 days)	7	12.3 [3.13-48.7]	263	3.07 [2.66-3.54]	274	2.29 [1.97-2.65]
Current use (31-182 days)	24	24.7 [6.51-93.6]	706	4.60 [4.12-5.13]	731	3.10 [2.73-3.52]
Current use (183-365 days)	8	45.5 [7.26-286]	356	11.6 [9.75-13.8]	380	5.10 [4.20-6.20]
Current use (>365 days)	9	68.2 [5.32-875]	771	40.1 [32.2-49.9]	825	8.80 [6.92-11.2]
Recent use (post-exposure period)	6	38.3 [7.67-191]	279	3.15 [2.73-3.62]	286	2.64 [2.27-3.07]

**In Mondriaan there are no subjects with multiple events, IRR- incidence rate ratios, CI- confidence interval, SCCS- self-controlled case series, THIN- The health improvement network, *Reference period comprises the no use (from start of the observation period until the first antidepressant prescription) and the past use (a period starting after the recent use until becoming a user again or the end of the study).

Table 5. Incidence rate ratios of hip/femur fracture based on a SCCS design when including only follow-up time after the first antidepressant prescription

	Mondriaan	
Exposure	Cases	IRR [95% CI] (Age adjusted)
Reference period*	28	1 (Ref)
Current use (1-30 days)	7	12.5 [5.39-29.0]
Current use (31-182 days)	24	11.1 [5.95-20.6]
Current use (183-365 days)	8	9.01 [3.46-23.5]
Current use (>365 days)	9	7.91 [1.80-34.8]
Recent use (post-exposure period)	6	6.24 [2.58-15.1]

IRR- incidence rate ratios, CI- confidence interval, SCCS- self-controlled case series, THIN- The health improvement network, *Reference period comprises the no use (from start of the observation period until the first antidepressant prescription) and the past use (a period starting after the recent use until becoming a user again or the end of the study).

Table 6 shows the results for the analysis in which subjects who were prescribed an antidepressant in the six months prior to the start of follow-up. In THIN, the IRRs were in line with the main analysis in Table 2. However, in Mondriaan, the IRRs were lower compared to the main analysis. For example, the IRR for the first 30 days of exposure was 1.59 [0.82-3.06] compared to IRR 3.22 [1.51-6.84] in the main analysis.

Table 6. Incidence rate ratios of hip/femur fracture based on a SCCS design including prevalent and new users

	Mondriaan		THIN	
Exposure	Cases	IRR [95% CI] (Age adjusted)	Cases	IRR [95% CI] (Age adjusted)
Reference period*	146	1 (Ref)	5278	1 (Ref)
Current use (1-30 days)	11	1.59 [0.82-3.06]	430	1.46 [1.32-1.62]
Current use (31-182 days)	49	1.82 [1.24-2.67]	1247	1.43 [1.33-1.53]
Current use (183-365 days)	19	1.35 [0.77-2.34]	712	1.38 [1.26-1.51]
Current use (>365 days)	56	1.90 [1.16-3.10]	1990	1.34 [1.23-1.46]
Recent use (post-exposure period)	12	1.45 [0.78-2.69]	436	1.35 [1.22-1.50]

IRR- incidence rate ratios, CI- confidence interval, SCCS- self-controlled case series, THIN- The health improvement network, *Reference period comprises the no use (from start of the observation period until the first antidepressant prescription) and the past use (a period starting after the recent use until becoming a user again or the end of the study).

Table 7 shows the results of the analysis in which exposure periods were redefined. Results were consistent with the main analysis. The highest risk of HF associated with antidepressant use was observed for the first 90 days of exposure, IRR 2.85 [1.69-4.80] and 1.58 [1.45-1.72] for Mondriaan and THIN, respectively.

Table 7. Incidence rate ratios of hip/femur fracture based on a SCCS design for different definition of exposure categories

	Mondriaan		THIN	
Exposure	Cases	IRR [95% CI] (Age adjusted)	Cases	IRR [95% CI] (Age adjusted)
Reference period*	82	1 (Ref)	4429	1 (Ref)
1-90 days	21	2.85 [1.69-4.80]	700	1.58 [1.45-1.72]
>90 days and recent use	33	2.27 [1.33-3.89]	1796	1.37 [1.27-1.48]

IRR- incidence rate ratios, CI- confidence interval, SCCS- self-controlled case series, THIN- The health improvement network, *Reference period comprises the no use (from start of the observation period until the first antidepressant prescription) and the past use (a period starting after the recent use until becoming a user again or the end of the study).

The effect of different classes of antidepressants is shown in Table 8. In Mondriaan, the IRR of HF was higher among SSRI users than among TCA users: IRR 4.39 [2.42-7.99] for SSRI and IRR 1.19 [0.58-2.41] for TCA. In THIN the association between SSRI and HF and the association between TCA and HF were similar.

Table 8. Incidence rate ratios of hip/femur fracture based on a SCCS design for classes of antidepressant drugs

	Mondriaan		THIN	
Exposure	Cases	IRR [95% CI] (Age adjusted)	Cases	IRR [95% CI] (Age adjusted)
Reference period*	82	1 (Ref)	4,429	1 (Ref)
Current SSRI use	35	4.39 [2.42-7.99]	1,303	1.49 [1.36-1.63]
Current TCA use	13	1.19 [0.58-2.41]	687	1.49 [1.33-1.66]
Current SSRI and TCA use	-	-	220	1.25 [1.00-1.55]
Recent use (post-exposure period)	6	2.23 [1.00-4.96]	286	1.37 [1.21-1.55]

IRR- incidence rate ratios, CI- confidence interval, SCCS- self-controlled case series, SSRI- selective serotonin reuptake inhibitor, TCA- tricyclic antidepressant, THIN- The health improvement network, *Reference period comprises the no use (from start of the observation period until the first antidepressant prescription) and the past use (a period starting after the recent use until becoming a user again or the end of the study).

To assess the impact of adjustment for the potential time-dependent confounding by benzodiazepine use, this variable was added to the age-adjusted model. Observed IRR were similar compared to the main analysis (Table 9).

Table 9. Incidence rate ratios of hip/femur fracture based on a SCCS design when possible time-dependent confounding by benzodiazepine use as well as age was adjusted for

	Mondriaan	
Exposure	Cases	IRR [95% CI]
Reference period*	82	1 (Ref)
Current use (1-30 days)	7	4.62 [2.35-9.11]
Current use (31-182 days)	24	2.98 [1.91-4.63]
Current use (183-365 days)	8	2.02 [0.98-4.16]
Current use (>365 days)	9	1.61 [0.83-3.11]
Recent use (post-exposure period)	6	2.38 [1.05-5.38]

IRR- incidence rate ratios, CI- confidence interval, SCCS- self-controlled case series, THIN- The health improvement network, *Reference period comprises the no use (from start of the observation period until the first antidepressant prescription) and the past use (a period starting after the recent use until becoming a user again or the end of the study).

DISCUSSION

This study shows that the SCCS design provides substantially biased results when subjects are censored at the event as well as starting follow-up with the first prescription during the study period. This phenomenon has been described before,[19,26] but our analyses using empirical data show that the impact is large and the effect estimates appear extremely biased. Moreover, when the analyses were focused on the possible event-exposure dependence, compared to the reference period, there appeared to be an increased risk during the 30 days prior to the start of antidepressant, suggesting event-exposure dependence. However, when the SCCS design were used correctly (our main analysis), the results of the SCCS design are in line with other study designs including the cohort, case-control and case-crossover design.[1–3] In this particular study, the SCCS results were consistent when using different definitions of risk periods or inclusion of subjects who were already exposed at the start of the study period. In addition, when we assessed the impact of adjustment for the potential time-dependent confounding by benzodiazepine use, the IRRs are in line with the main analysis, indicating the benzodiazepine may not be a strong confounder in our study.

When subjects are censored at their first HF, the assumption underlying the SCCS design that events can be recurrent is violated, the IRRs are much larger than those of IRRs including all events suggesting substantial biased estimates even after adjustment for age. Also, when subjects are censored at their last HF, the IRRs appear substantially biased. Although in the

latter setting the events are recurrent, the bias may be due to the follow-up time that is differentially removed from the analysis (i.e., particularly exposed observation time is removed from the analysis, suggesting that event may affect the probability of exposure). In our study this bias is upward because the exposed observation time that is removed is much larger than the unexposed observation time. Indeed, Weldeselassie et al[19] and Farrington et al[27] stated that censoring at the event produces a biased estimate that is unpredictable in direction. Hence, the recurrent HF may not be independent in our analysis.[17] In addition, this finding points out that the observation periods may be event dependent, consequently another key assumption (the occurrence of the event must not affect the observation period) may be violated here.[19,33] Farrington and Whitaker[34] argued that the SCCS is often robust against violations of this assumption, which cannot be supported by the findings in our study.

We tested the assumption that the occurrence of an HF must not alter the probability of subsequent antidepressants prescription using a pre-exposure category of 30 days prior to initiation of antidepressant therapy. In both databases, the event rates during this period are different from the event rates in the reference period, suggesting that this assumption may be violated.[17]

When analysis is done including only follow-up time after the first antidepressant prescription, the IRRs are larger than those of IRRs including complete follow-up time suggesting biased estimates. This bias may be due to the magnitude of follow-up time that is removed from the analysis which is higher in reference period compared to the current periods. Moreover, some authors argued that this bias is due to cases that are selectively excluded from the analysis.[26,35]

The SCCS yields similar results across the two databases regarding different definitions of risk windows. Moreover, when analysis is done including the subjects who were prescribed antidepressants in the 6 months before the start of follow-up, the pattern of IRR is similar to our main analysis (Table 6 vs 2) mainly in the THIN. However, the IRRs for Mondriaan are lower than those of IRRs in Table 2 because the duration of follow-up time in the current periods increases more (about 45%) than the reference period.

Our study has several strengths. First, this is the first SCCS study across databases from different countries. Second, the definitions of the observation periods and risk windows in the two databases are based on a common study protocol. Third, we performed extensive analyses to assess the impact of violations of the key assumptions and different definitions of risk windows. Limitations of our study are that we used information on prescribed antidepressants, rather than those dispensed by the pharmacy. Moreover, by definition a subject could not experience an event during the 12 months following an event, which may induce an immortal time bias.[36] Furthermore, a direct comparison of the effects of SSRI and TCA is impaired by possible incomparability of SSRI and TCA users, except for patients who switch between these drugs.

In conclusion, different definitions of risk windows or inclusion of subjects who were already exposed at the start of the study period did not have a major impact on the results from the SCCS. However, the design is sensitive and produces apparent biased results in both databases when the follow-up time is censored at the event or follow-up time starts with the first use of antidepressants. The performance of this design may differ across studies and across databases. Therefore, in each SCCS study, correct specification of the SCCS design should be carefully assessed and reported.

REFERENCES

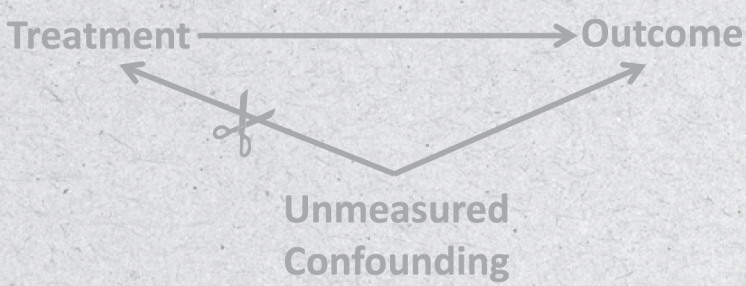
1. Van Den Brand MWM, Samson MM, Pouwels S, et al. Use of anti-depressants and the risk of fracture of the hip or femur. *Osteoporosis Int* 2009; 20(10): 1705-13.
2. Thapa PB, Gideon P, Cost TW, Milam AB, Ray WA. Antidepressants and the risk of falls among nursing home residents. *N Engl J Med* 1998; 339(13): 875-82.
3. Ginzburg R, Rosero E. Risk of fractures with selective serotonin-reuptake inhibitors or tricyclic antidepressants. *Ann Pharmacother* 2009; 43(1): 98-103.
4. Hubbard R, Farrington P, Smith C, Smeeth L, Tattersfield A. Exposure to tricyclic and selective serotonin reuptake inhibitor antidepressants and the risk of hip fracture. *Am J Epidemiol* 2003; 158(1): 77-84.
5. Pratt N, Roughead EE, Ramsay E, Salter A, Ryan P. Risk of hospitalization for hip fracture and pneumonia associated with antipsychotic prescribing in the elderly: A self-controlled case-series analysis in an Australian health care claims database. *Drug Safety* 2011; 34(7): 567-75.
6. Douglas IJ, Smeeth L. Exposure to antipsychotics and risk of stroke: self controlled case series study. *BMJ* 2008; 337:a1227.
7. Douglas IJ, Langham J, Bhaskaran K, Brauer R, Smeeth L. Orlistat and the risk of acute liver injury: Self controlled case series study in UK Clinical Practice Research Datalink. *BMJ (Online)* 2013; 346:f1936(7906).
8. Douglas IJ, Evans SJ, Pocock S, Smeeth L. The risk of fractures associated with thiazolidinediones: A self-controlled case-series study. *PLoS Medicine* 2009; 6:e1000154. doi:10.1371/journal.pmed.1000154(9).
9. Grosso A, Douglas I, Hingorani A, MacAllister R, Smeeth L. Oral bisphosphonates and risk of atrial fibrillation and flutter in women: A self-controlled case-series safety analysis. *PLoS ONE* 2009; 4:e4720. doi:10.1371/journal.pone.0004720(3).
10. Gribbin J, Hubbard R, Gladman J, Smith C, Lewis S. Risk of falls associated with antihypertensive medication: Self-controlled case series. *Pharmacoepidemiol Drug Saf* 2011; 20(8): 879-84.
11. Wilson K, Hawken S. Drug safety studies and measures of effect using the self-controlled case series design. *Pharmacoepidemiol Drug Saf* 2013; 22(1): 108-10.
12. Wijlaars LPMM, Nazareth I, Whitaker HJ, Evans SJW, Petersen I. Suicide-related events in young people following prescription of SSRIs and other antidepressants: A self-controlled case series analysis. *BMJ Open* 2013; 3:e003247(9). doi:10.1136/bmjopen-2013-003247.
13. Ramsay EN, Pratt NL, Ryan P, Roughead EE. Proton pump inhibitors and the risk of pneumonia: A comparison of cohort and self-controlled case series designs. *BMC Medical Research Methodology* 2013; 13:82,doi:10.1186/1471-2288-13-82(1). doi:10.1186/1471-2288-13-82.
14. Farrington CP. Relative incidence estimation from case series for vaccine safety evaluation. *Biometrics* 1995; 51(1): 228-35.

15. Whitaker HJ, Farrington CP, Spiessens B, Musonda P. Tutorial in biostatistics: The self-controlled case series method. *Stat Med* 2006; 25(10): 1768-97.
16. Hallas J, Pottegård A. Use of self-controlled designs in pharmacoepidemiology. *J Intern Med* 2014; 275(6): 581-9. doi:doi: 10.1111/joim.12186.
17. Whitaker HJ, Hocine MN, Farrington CP. The methodology of self-controlled case series studies. *Stat Methods Med Res* 2009; 18(1): 7-26. doi:10.1177/0962280208092342 [doi].
18. Whitaker H. The self controlled case series method. *BMJ* 2008; 337:a1069.
19. Weldeselassie YG, Whitaker HJ, Farrington CP. Use of the self-controlled case-series method in vaccine safety studies: Review and recommendations for best practice. *Epidemiol Infect* 2011; 139(12): 1805-17.
20. Arnheim-Dahlstrom L, Hallgren J, Weibull CE, Sparen P. Risk of presentation to hospital with epileptic seizures after vaccination with monovalent AS03 adjuvanted pandemic A/H1N1 2009 influenza vaccine (Pandemrix): self controlled case series study. *BMJ* 2012; 345: e7594. doi:10.1136/bmj.e7594 [doi].
21. Huang W, Yang H, Liao T, Wu W, Yang S, Chih Y, Chuang J. Safety of pandemic (H1N1) 2009 monovalent vaccines in Taiwan: a self-controlled case series study. *PloS one* 2013; 8(3): e58827.
22. Romio S, Weibel D, Dieleman JP, et al. Guillain-Barré syndrome and adjuvanted pandemic influenza A (H1N1) 2009 vaccines: a multinational self-controlled case series in Europe. *PloS one* 2014; 9(1): e82222.
23. Escolano S, Hill C, Tubert-Bitter P. A new self-controlled case series method for analyzing spontaneous reports of adverse events after vaccination. *Am J Epidemiol* 2013; 178(9): 1496-504.
24. Xu S, Newcomer S, Nelson J, et al. Signal detection of adverse events with imperfect confirmation rates in vaccine safety studies using self-controlled case series design. *Biometrical Journal* 2014; 56(3): 513-25.
25. Langan SM, Minassian C, Smeeth L, Thomas SL. Risk of Stroke Following Herpes Zoster: A Self-Controlled Case-Series Study. *Clinical Infectious Diseases* 2014; : ciu098.
26. Hua W, Sun G, Dodd CN, et al. A simulation study to compare three self-controlled case series approaches: Correction for violation of assumption and evaluation of bias. *Pharmacoepidemiol Drug Saf* 2013; 22(8): 819-25.
27. Paddy Farrington C, Anaya-Izquierdo K, Whitaker HJ, Hocine MN, Douglas I, Smeeth L. Self-controlled case series analysis with event-dependent observation periods. *Journal of the American Statistical Association* 2011; 106(494): 417-26.
28. Farrington CP, Whitaker HJ, Hocine MN. Case series analysis for censored, perturbed, or curtailed post-event exposures. *Biostatistics* 2009; 10(1): 3-16.
29. Abbing-Karahagopian V, Huerta C, Souverein P, et al. Antidepressant prescribing in five European countries: application of common definitions to assess the prevalence, clinical observations, and methodological implications. *Eur J Clin Pharmacol* 2014; 70(7): 849-57.
30. ENCePP-. Use of antidepressants and risk of hip/femur fracture. A methodological comparison across data sources and epidemiological design. Electronic Register of Studies, 2014. Available from: http://www.encepp.eu/encepp/openAttachment/fullProtocolLatest/5847;jsessionid=mnbHT9QfbfCJSQLnPl91LQkW3MhQ61rhRXG9gtx23tT8hvCLzcDF!-1931278576 (Date of access: 16 July 2014).
31. Gardarsdottir H, Souverein PC, Egberts TCG, Heerdink ER. Construction of drug treatment episodes from drug-dispensing histories is influenced by the gap length. *J Clin Epidemiol* 2010; 63(4): 422-7.
32. Simpson SE. A Positive Event Dependence Model for Self-Controlled Case Series with Applications in Postmarketing Surveillance. *Biometrics* 2013; 69(1): 128-36.

33. Farrington P. Censoring on outcome is not valid in self-controlled case series studies. *J Clin Epidemiol* 2013; 66(12): 1428-9.
34. Farrington CP, Whitaker HJ. Semiparametric analysis of case series data. *Journal of the Royal Statistical Society.Series C: Applied Statistics* 2006; 55(5): 553-94.
35. Maclure M, Fireman B, Nelson JC, Hua W, Shoaibi A, Paredes A, Madigan D. When should case-only designs be used for safety monitoring of medical products? *Pharmacoepidemiol Drug Saf* 2012; 21(SUPPL. 1): 50-61.
36. Suissa S. Immortal time bias in pharmacoepidemiology. *Am J Epidemiol* 2008; 167(4): 492-9.

Unmeasured confounding IV PERR SCCS Observational studies
PROTECT Causal Inference Statistical methods পরিসংখ্যানগত পদ্ধতি
Statistische methoden IV PERR SCCS Pseudo Randomization
Pharmacoepidemiology IV Analyses PERR SCCS
General Practice Databases Residual Confounding
Electronic Healthcare Databases Epidemiology
Time-fixed Confounding Time-fixed Exposure
Time-varying Exposure
Time-varying Confounding
Non-randomized Studies
IV PERR SCCS
PROTECT 2SLS
IV IV

CHAPTER VI
GENERAL DISCUSSION



Randomized controlled trials may not always be feasible to estimate the effect of pharmacological treatments, for instance, to detect rare adverse drug reactions. In that case, observational studies may be a viable alternative and such studies may thus provide important evidence about the comparative safety and effectiveness of treatments.[1–7] However, in observational studies a randomized assignment of treatments is absent; hence, the treatment effect can be biased due to possible confounding by patients' characteristics.[4,8–10] A confounder is a prognostic factor of the outcome which is associated with the exposure and causes a biased estimate of the exposure effect on the outcome.[11,12] Several statistical methods, e.g., multivariable regression, propensity score analysis, have been used to control for confounding by measured confounders.[7,8,10,13,14] However, confounding by unmeasured patient characteristics (e.g., activities of daily living, disease severity)[15] is also a potential problem of observational studies.[8,10] Additional methods have been used to control for unmeasured confounding in observational pharmacoepidemiological studies.[7,10,15–23] Among the novel methods, in this thesis, we aimed to evaluate the performance of instrumental variable (IV) analysis, prior event rate ratio (PERR) adjustment, and the self-controlled case series (SCCS) design. We evaluated these methods in simulation studies and in empirical studies based on electronic healthcare records databases from different countries using two case examples: a study of the effect of inhaled long-acting beta2-agonist (LABA) use on myocardial infarction and a study of the effect of antidepressant use on hip fracture. In this chapter, we discuss the findings of our studies and its limitations, compare our results with existing literature, and point out possible strengths and limitations of these methods. Finally, we provide recommendations for proper application of the methods in pharmacoepidemiology. We will use the terms "treatment" and "exposure" interchangeably. Also, we will use the term "IV estimate" to indicate the exposure effect obtained by IV analysis.

INSTRUMENTAL VARIABLE ANALYSIS

IV analysis is becoming increasingly popular to control for unmeasured confounding in observational pharmacoepidemiological research.[24–26] An IV requires three basic assumptions: 1) the IV is associated with treatment; 2) the IV affects the outcome only through the treatment; and 3) the IV is independent of confounders.[18,27–29]

Instrumental variable analysis in cohort and (nested) case-control designs

The performance of IV analysis in a cohort design is relatively well known in pharmacoepidemiological studies.[18,30–34] IV analysis can also be applied in a nested case-control (NCC) study, which is nested within a larger cohort of known size. In Chapter 3.1, we showed that the selection and construction of IVs in the NCC study are in line with a cohort study if the control subjects are sampled appropriately. Moreover, like cohort studies, the IV estimates in NCC studies performed poorly in case of weak IV-exposure associations, particularly for small sample sizes. The variability of the estimates is more pronounced for rare outcomes especially in NCC studies in which the effective sample size is smaller than

in the cohort studies. This can be partly remedied by increasing the sample size, or by increasing the number of controls per case.[Chapter 3.1]

Assessment of the assumptions of IV analysis

The first IV assumption is empirically verifiable by using available methods, e.g., partial F-statistic,[35] correlation,[36] or odds ratio.[34,37] The F-statistic from the first-stage IV model is widely used,[33,38,39] and there is a rule of thumb that if the value is less than 10, the IV is weakly associated with exposure.[5,27,35,40,41] Nevertheless, the IV estimates may still be biased when the values of the F-statistic are around 10.[36,Chapter 3.1] Moreover, the F-statistic strongly depends on the sample size.[Chapter 3.1] Alternatively, other methods (e.g., correlations or odds ratio) can be used to identify whether an IV is weak or strong. We found that a weak IV can be identified when the correlation between IV and exposure is less than 0.15 for continuous or the odds ratio is less than 2.0 for binary IV and exposure; although these specific cut-off values may vary between studies.[Chapter 3.1]

When the IV is weak, the exposure effect estimates are often inconsistent and biased with wide confidence intervals even if all other IV assumptions are met.[27,42–48] In our empirical studies (Chapters 4.1 and 4.2), the IV estimates were uninformative in case of weak IVs, despite large samples sizes.[43] For example in Chapter 4.1, the exposure effect with a weak IV (physician prescribing preference, PPP, using single previous prescription of a physician in the asthma cohort) is highly variable with wide 95% confidence intervals; unadjusted and adjusted hazard ratio: 35.6 [3.86-284] and 7.41 [0.77-78.2], respectively. A weak IV may also amplify bias due to small violations of the second and third assumptions.[29,Chapter 3.1] Consequently, IV estimates may be more biased than estimates obtained by conventional methods.[29,36,49]

The second and third IV assumptions are unverifiable or not directly testable as they involve unobservable variables.[27,29,50] However, several approaches[5,51–53] were proposed to assess the validity of these assumptions. Several other authors[18,30,37,54,55] provided supportive evidence for the validity of the third assumption, by describing the balance of observed patient characteristics between IV categories. Alternatively, an imbalance in observed patient characteristics may falsify this assumption.[49] In Chapter 3.2, we proposed the standardized difference, a robust balance measure used in propensity scores analysis,[56,57] to falsify the third assumption by checking independence between an IV and measured confounders. If measured confounders are insufficiently balanced between IV categories, indicated by standardized difference values deviating from zero (e.g., > 0.10),[56] this may also imply imbalance of unmeasured confounders, even after conditioning on measured confounders. In that case, the third assumption is violated. However, Angrist and Piskhe[40] argued that when there may still be imbalance in confounders' distributions across IV levels, the third assumption is more likely to hold after conditioning on measured confounders. Baiocchi et al.[5] also stated that if a measured confounder(s) is only a proxy for a true confounder then an imbalance of measured confounder(s) may induce an association between the IV and the

unmeasured part of the true confounder. On the other hand, if measured confounders are balanced, investigators should rely on background knowledge to argue that such balance could be carried over to unmeasured confounders.[18,30]

The second IV assumption, that the IV affects the outcome only through the exposure, cannot be verified from the data. Theoretical justification by using clinical knowledge has been used to argue whether this assumption is fulfilled or not.[18,33,58] For instance, in Chapter 4.1, this assumption may be violated because the IV may be associated with concomitant medication use (e.g., inhaled corticosteroid) of the exposure LABA as there is evidence that inhaled corticosteroid use influences the outcome (myocardial infarction).[59,60] In that case, the assumption two and three seem similar, however, Swanson and Hernan[49] mentioned that it is essential to assume them separately and clinical knowledge is always important about their validity in a particular study.

If the above mentioned basic assumptions are met, IV analysis allows to estimate lower and upper bounds, but not a point estimate, for the average causal effect (ACE) of an exposure on an outcome.[5,29,49] However, one may be interested in the point estimate of the exposure effect (e.g., ACE or complier average causal effect). For that, additional assumptions are required, i.e., homogeneous exposure effects (exposure effects are constant across the study population), or monotonicity (there are no defiers), or no effect modification by the IV.[5,29,49,61] However, these additional IV assumptions are hard to verify empirically and require theoretical justification.[29,49]

Methods for instrumental variable estimation

IV analysis generally involves a two-stage modelling approach. In the first stage, the effect of the IV on exposure is estimated, whereas in the second stage, outcomes are compared in terms of predicted exposure rather than the actual exposure.[62] Several methods such as two-stage least squares (2SLS), two-stage residual inclusion, generalized method of moments, structural mean models, are available for different types of IV, exposure and outcome and different methods may be applicable, each of which have advantages and disadvantages (summarized in Tables 1 and 2 in Chapter 2.2). The 2SLS method has been widely used in pharmacoepidemiology.[25,50,63–68] Although several authors[5,29,69] argued that IV analysis using nonlinear models (e.g., logistic regression model, Cox model) may not consistently estimate the exposure effects, several studies[70–72] applied IV analysis using the nonlinear models (e.g., Cox model) because it is much easier for a clinician to understand the exposure effects.

If the assumptions of IV analysis can be reasonably assumed to hold, an unadjusted IV estimate may provide a valid exposure effect, although adjusted IV analysis may be more precise.[38,40,73] If there is any evidence that IV assumptions are violated, the IV should be treated as invalid and one should not proceed with IV analysis even after conditioning on measured confounders in the IV models because several authors[27,74,Chapter 3.2] already reported that violation of any IV assumption can magnify the bias due to confounding. However,

imbalance of measured confounders can be controlled for in IV analysis, if the unmeasured confounders can be assumed to be conditionally independent of the IV.[Chapter 3.2]

Instrumental variable analysis across the electronic healthcare databases

Electronic healthcare record databases often have limited or inaccurate information on some potential confounders, such as alcohol consumption,[17,18] and thus IV analysis is increasingly popular in database studies.[24,26,75] Exposure effects may differ considerably depending on the definitions of IVs.[18,33,36,53] The IV PPP is most commonly used in pharmacoepidemiology.[18,31,33,39,58] This IV can be defined in different ways, for example, based on the proportion of prescriptions of a particular drug by a certain physician.[18,33,36,53] In Chapter 4.1 and 4.2, we evaluated the IV analyses with different definitions of PPPs across multiple databases using both time-fixed and time-varying exposures.

Among the different IVs, the IV PPP based on the previous 10 prescriptions appears to perform better than the other considered IVs, with respect to strength of the association between IV and exposure and balance of confounders between IV categories.[Chapter 4.1,53,58] However, the optimal number of prescriptions included in the IV may differ between studies. For instance in Chapter 4.2, with respect to our assessment criteria (e.g., correlation ≥ 0.15, or odds ratio ≥ 2.0, standardized difference < 0.10), we did not identify any IV that appears consistently valid across the databases. Obviously, choices on the number of prescriptions used for the IV should depend on the IV assumptions, not on the IV effect estimate.[Chapters 4.1 and 4.2]

In case of time-varying exposures, the IVs were defined as the proportions of time for LABA or SSRI use for a practice. In all databases, neither of the IVs we considered appears to be valid. The main limitation is the weak association between the IVs and the exposure, due to limited variation in prescribing patterns between practices.

In addition, there appears to be a trade-off between the strength of the IV and the balance of confounders across IV strata.[27] In all databases we found that the IV PPP based on the last single prescription is weakly associated with the exposure but confounders were balanced across the IV categories. On the other hand, in the Mondriaan database, IVs were strongly associated with the exposure, but measured confounders were imbalanced across IVs categories.[Chapter 4.1 and 4.2]

In the analysis of the pooled data, the variation of preferences between practices increased (compared to the analysis per database), resulting in a stronger IV and hence, IV estimates were more precise.[Chapter 4.1 and 4.2] Moreover, as IV analysis produces unstable estimates when the outcome is rare,[76, Chapter 3.1] pooling databases may overcome this limitation.[Chapter 4.1 and 4.2]

Researchers generally choose IVs for a particular study based on existing literature or expert knowledge. However, if an IV appears to be valid in one study, this cannot be generalized to other studies. Therefore, a clear objective strategy to evaluate the (apparent) validity of

possible IVs should be included in the evaluation process of every IV analysis. Although the PPP is commonly used for pharmacoepidemiological studies, it is not the only possible IV. Differential implementation of guidelines between (similar) regions, or evaluating the implementation of guidelines (before-after comparison) may provide valid IVs. We propose that new IVs are considered that allow for estimating unbiased treatment effects of safety and effectiveness in pharmacoepidemiology.

In summary, we observed that considering several plausible IVs in a study may be useful to evade the uncertainty of the IV estimates based on a single IV. The performance of IV analyses varied between time-fixed and time-varying exposures, across the databases and strongly depends on the definition of IVs and size of the samples.[Chapter 4.1 and 4.2] Our multiple databases studies show that it is still challenging to obtain a valid IV in pharmacoepidemiological studies especially for studies of time-varying exposures.

PRIOR EVENT RATE RATIO ADJUSTMENT METHOD

The PERR adjustment method has been used to control for time-fixed confounding (measured and unmeasured).[19,77–80] The exposure effect is estimated by the ratio of two rate ratios (RRs): RR after initiation of exposure (RR in post period) and the RR prior to initiation of exposure (RR in prior period).[16,78] It requires assumptions about constant confounding effects across prior and post periods, absence of confounder-by-treatment interaction, and non-lethal events as outcome.[7,78,80] In Chapter 5.1, we assessed this method using simulated data to clarify constraints and understand its proper applicability.

When confounding effects are not constant across the prior and post periods, the PERR method yields biased treatment effect estimates (e.g., Figure 2 in Chapter 5.1). Moreover, when prior events influence the probability of subsequent exposure, or when prior events directly influence the probability of events in the post period, or when there is considerable dropout in the post period, the method also shows biased results. In particular situations, e.g., when events in the prior period strongly influence the probability of exposure, the amount of bias can be larger than that found in conventional analysis. In addition, the variation of the estimates is larger for the PERR method than for the conventional method. The variation decreases as incidence rates increase in both periods.[78,Chapter 5.1]

SELF-CONTROLLED CASE SERIES DESIGN

The SCCS design uses information on case subjects only and controls for time-fixed measured and unmeasured confounding.[81,82] It requires three key assumptions: i) events are recurrent and independent or are unique and uncommon; ii) the occurrence of an event must not alter the probability of subsequent exposure; and iii) the occurrence of an event should not censor the observation period.[22,81–83] The incidence rate ratio is defined as the rate of events during exposed periods divided by the rate during all other unexposed observation periods.[22,82,84,85] While several pharmacoepidemiological studies have been

conducted using this design,[85–89] in Chapter 5.2, we evaluated the impact of violations of its assumptions and different definitions of observation periods and risk windows in two databases (Mondriaan, Netherlands and THIN, UK) in a study of antidepressants use and risk of hip fracture.

When the SCCS design is applied correctly, the exposure effects are stable and more efficient than those of an ordinary cohort design (Table 2 in Chapter 5.2). However, the SCCS design is sensitive to violation of the assumptions and provides biased results particularly when subjects are censored at their event.[Chapter 5.2,83,90] Our SCCS analyses across different databases show that the impact of violations of the assumptions can be large, in which case the effect estimates appear extremely biased. For example in the THIN database, incidence rate ratios for > 365 days of exposure period are 1.26 [95% confidence interval 1.13-1.42] when subjects are followed until end of the study and 40.1 [32.2-49.9] when censoring is at the first event.[Chapter 5.2] We found that when the analysis is restricted to those subjects whose follow-up starts with the initiation of the exposure (i.e., incident users), the SCCS appears to produce biased results as compared to the analysis in which subjects are followed from the beginning to the end of the predefined study period. This bias may be due to the follow-up time that is differentially removed from the analysis or cases that are selectively excluded from the analysis.[21,91]

POSSIBLE LIMITATIONS AND FUTURE RESEARCH

Several possible limitations of our studies need to be acknowledged. The limitations may provide insights for future research.

In our simulation studies of IV analysis, we did not assess the bias and variability of IV estimates under heterogeneous exposure effects, time-varying exposure and confounding. In pharmacoepidemiology, however, exposures generally involve medications, which are more often time-varying with possibly heterogeneous effects on the study population. Hence, it would be worthwhile to assess of bias and variability of effect estimates in these settings. Moreover, IV methodology is not well developed for time-varying exposure with survival outcomes in which censoring is a common complication.[5,92,93] Although, in that case, Robins' g-estimation of structural nested models have been proposed,[29,94] these have not been commonly used in pharmacoepidemiology. Alternatively, we used Cox model (second-stage IV model) in our empirical IV analysis, though some authors argued that this model is not theoretically motivated and may not yield estimates that are causally interpretable.[38] Therefore, it is necessary to further evaluate this method or to develop an alternative method that can easily be applied in these settings.

When more than two study drugs are included in an exposure, the preference of physician may not be measured correctly. Hence, in Chapter 4.1, the IV PPP may reflect an incorrect preference of a physician as we compared LABA versus non-LABA where non-LABA included short-acting beta-2-receptor agonist (SABA), short- and/or long-acting muscarinic

antagonist. This indicates that the IV PPP may not be suitable for our LABA example, and thus other IVs should be considered for further studies.

Although we applied a robust balance measure (standardized difference) to falsify the third IV assumption in our empirical studies,[57,95] this could fail to falsify an IV even when the third assumption is violated.[49] For example in Chapter 4.2, the IV PPP based on the last 10 prescriptions appeared to be valid for the analysis of time-fixed SSRI exposure, however, the unadjusted and adjusted IV estimates were very different, which suggests an association between the IV and measured confounders.[38]

To define a time-varying IV (e.g., the proportion of time of LABA use (Chapter 4.1) or the proportion of time for SSRI use (Chapter 4.2)) that is easy to apply, we considered exposure as two categories (i.e., current SSRI versus current or recent or past TCA, or recent or past SSRI), which may not be a perfect comparison groups. Moreover, with respect to our assessment criteria, these time-varying IVs violated at least one of the IV assumptions across the databases in both case examples and are therefore invalid. Hence, it is essential to create alternative IVs for time-varying settings.

We assumed that unmeasured confounding is a problem in our empirical studies. However, we did not conduct any sensitivity analysis to quantitatively assess the impact of unmeasured confounding.[96] Moreover, several potential unmeasured confounding factors (e.g., smoking, activities of daily living) are already known for part of the subjects included in our studies, and thus propensity score calibration may be an alternative option to control for these confounders.[97,98]

In our empirical studies, we only focused on confounding, however, other biases, e.g., information or selection bias, may influence the validity of the results. For example, for all empirical studies, only information on prescriptions was available; but, we do not know whether patients actually collected their medicines from the pharmacy and took them as prescribed. Furthermore, although confounding and effect modification (or interaction) are different issues, effect modification may induce problems of comparability of the study results across databases. For instance, in Chapter 4.2, the effect of SSRI on the risk of hip fracture may be modified by age.

In the assessment of the PERR method, we assumed incidence rates that were constant across prior and post periods. It would be of interest to extend the assessment to explore settings of different incidence rates in both periods.

Finally, although we used many realistic pharmacoepidemiological settings in our simulation studies, the results from simulations obviously do not reflect all possible scenarios. Extensive evaluation of methods using series of empirical studies (e.g., our IV analyses in different cohorts with multiple databases) may reveal the actual performance of methods to control for unmeasured confounding.

RECOMMENDATIONS FOR CHOOSING BETWEEN DIFFERENT METHODS TO CONTROL FOR UNMEASURED CONFOUNDING

In this section, we provide recommendations for choosing between methods to control for unmeasured confounding. In Table 1, the characteristics and assumptions of different methods to control for unmeasured confounding are summarized.

Before applying any method to control for unmeasured confounding, we recommend to use theoretical motivations combined with clinical knowledge and empirical evidence to motivate that unmeasured confounding is an important threat to the validity of a particular study. For example, when studies involve intended drug effects, confounding by indication may induce a large amount of bias in the exposure effects (i.e., there is a big concern about unmeasured confounding).[2,5,99] In that case, IV analysis for example may play an important role provided that the IV is available and satisfies the IV assumptions. On the other hand, in studies of adverse drugs effects in which individual risks of particular outcomes (e.g., immunological adverse effects)[2] is often unknown and not influencing the prescription of drugs,[2] the impact of unmeasured confounding may be small and thus ordinary methods to control for confounding may be applied to estimate exposure effects.[100] Nevertheless, when an adverse effect is well known (e.g., relation between aspirin use and risk of gastrointestinal bleeding),[99,100] the risk is often taken into account when prescribing and hence, strong confounding by contraindication may be present and the methods that make the assumption of no unmeasured confounding may yield biased results.

If there is a strong concern about unmeasured confounding in a study, it may be worthwhile to replace the assumption of no unmeasured confounding by perhaps more plausible assumptions required for IV analysis, the PERR method, or the SCCS design. Importantly, some of these assumptions are not verifiable and expert knowledge is always necessary to justify them.

The type of data that are available for a pharmacoepidemiologic study may guide the choice for a method to handle unmeasured confounding. For instance, IV analysis can be applied in a cohort study as well as in a nested case-control (NCC) study. However, when applying IV analysis in the NCC design, we recommend increasing the number of controls per case (at least 4 controls per case) in order to increase effective sample size and reduce variability of the IV estimates. Additionally, as the NCC design does not use the entire cohort, this design requires a stronger IV-exposure association than the cohort design.[Chapter 3.1] The PERR method, requires that data are available on two periods: i.e., before and after initiation of the exposure. It cannot be applied when patients have no records prior to the study start.[19,77] Also, an inappropriate duration of the prior time period may influence the validity of the estimates from the PERR analysis.[78] Finally, the SCCS design uses information of cases only. Nevertheless, incorrect definitions of observation period may seriously bias SCCS results. Hence, the possibility to apply any of these methods depends on the research setting.

The choice for a particular method to handle unmeasured confounding may also depend on the outcome of interest in a particular study. Theoretically, any type of outcome can be studied using the IV method. However, the PERR method is outcome specific, meaning that its validity cannot be extrapolated from one outcome to another (e.g., it cannot be applied for the lethal outcomes).[78] The SCCS method is also not applicable for all types of outcomes (e.g., if occurrence of events increases mortality or if the outcome is mortality). In addition, one of the conditions of the PERR and the SCCS is that events are generally recurrent and independent. However, the SCCS method can also be applied if the events are unique and very uncommon. In addition, if all events are to happen at exactly the same time/age in the SCCS design it is impossible to estimate the exposure effects, which is also the case when applying the PERR method if the number of events is too small in the prior period compared to the post period in the PERR method.

The IV and the PERR methods yield biased and unstable estimates when studies involve small samples with rare events, but this problem is likely to be small in the SCCS design given that at least 2.5 events are expected in the risk periods.[83,101]

The PERR and SCCS methods may not be appropriate when events influence the probability of subsequent exposures, which is not the case with the IV method. For example, a study of statin use and risk of myocardial infarction or stroke would likely be biased when applying these methods. Because when a patient first experienced a myocardial infarction or stroke, (s)he has to start treatment (e.g., statin) in order to avoid a secondary myocardial infarction or stroke, and consequently the health characteristics of this patient (e.g., cholesterol levels) are influenced by such type of therapy. Hence, the probability of the exposure (start statin therapy) is definitely influenced by previous outcome as well as the confounders. However, they may be valid methods in a situation where the probability of exposure and probability of events are independent. An example may be a study of the effect of proton-pump inhibitors on the risk of pneumonia or statin use and risk of hip fracture or lower urinary tract symptoms. However, if the rate of mortality or dropout is high, or the unmeasured confounder effects vary temporally, they may not provide unbiased results. Additionally, the PERR or the standard SCCS methods may produce biased estimates when the exposure is partially or completely contraindicated, which is also stressed by others.[91] An example of likely bias would be a study of antipsychotics use and the risk of stroke.

IV analysis suffers from a key shortcoming, which is that it is hard to find a valid IV in pharmacoepidemiological studies, in particular when exposure and confounders are time-varying. Although the PERR and SCCS methods can be suitable to study time-varying exposures, they may provide biased results when confounders change over time.[81,Chapters 5.1 and 5.2] However, measured time-varying confounding (e.g., by age) can be adjusted for in the PERR methods and the SCCS design.

The exposure effects from the IV and conventional analyses are not always directly comparable due to the underlying point identifying assumptions of the IV analysis.[102,Chapter

Table 1. Characteristics and assumptions of different methods to control for unmeasured confounding

	IV analysis	PERR adjustment	SCCS design
Basic principles and assumptions	- uses information of all subjects; design can be cohort or NCC - assumption: the IV is associated with exposure, affects the outcome only through the exposure, and is independent of confounders	- uses information of all subjects in two periods (prior and post initiation of the exposure); cohort design - assumption: constant confounding effects between prior and post periods, absence of confounder-by-treatment interaction	- uses only information of cases; design like a case-crossover, subjects act as their own control - assumption: events are recurrent and independent or are unique and uncommon over the study periods; the event must not alter the probability of exposure; and the event should not censor the observation period
Exposure status and study follow-up	- exposure: time-fixed or time-varying - censoring at first event can be handled	- exposure: time-fixed or time-varying - subjects should not be exposed in the prior period - censoring at first event in the prior period is not allowed - biased if outcome influences subsequent exposure	- exposure: time-varying but can be time-fixed - subjects can exposed before or after event - the follow-up cannot end with event; censoring at first event induces bias - biased if outcome influences the subsequent exposure - if all events are to happen at exactly the same time/age, then the method fails.
Outcome	- any type of outcome	- outcome specific (cannot be applied for lethal outcome), recurrent outcome	- not applicable for all outcomes (e.g., if occurrence of events increase mortality, or outcome is mortality)
Confounding (measured and unmeasured)	- perform better for time-fixed confounding, also applicable for time-varying confounding - measured confounders can be adjusted in the IV models	- only control for time-fixed confounding, bias when unmeasured confounding is time-varying - measured confounders can be adjusted	- controls for time-fixed confounding - some time-varying measured confounders can be adjusted in the model (e.g., age)
Sample size	- in finite samples, IV estimates may be biased in the same direction as conventional estimates	- for very rare events, it requires large sample size to estimate the exposure effect	- requires at least 2.5 events (in expectation) during the risk period
Data analytical method	- two-stage models (e.g., 2SLS)	- log-linear or Cox hazard model	- conditional Poisson regression

Table 1. Characteristics and assumptions of different methods to control for unmeasured confounding (*Continued*)

	IV analysis	PERR adjustment	SCCS design
Estimate	- RD/RR/OR/HR: from the second-stage IV model - provides different causal parameters, ACE, CACE, and ATT, depending on assumptions	- IRR or HR: ratio of two RRs: RR in post period and RR in prior period - ACE	- IRR: rate of events during exposed periods divided by the rate during all other observed periods - ACE
Standard error	- robust or bootstrap method - often larger than with conventional method	- bootstrap method - often larger than with conventional method	- robust method - often more efficient than cohort design
Application in pharmacoepidemiology	- challenge to identify IV	- has been applied in several studies (e.g., statins and myocardial infarction)	- has been applied in several studies and results appear to be valid

IV: instrumental variable, PERR: prior event rate ratio, SCCS: self-controlled case series, NCC: nested case-control, RD: risk difference, RR: risk ratio/rate ratio, IRR: incidence rate ratio, HR: hazard ratio, OR: odds ratio, ACE: average causal effect, CACE: complier average causal effect, ATT: average treatment effect for the treated individuals, 2SLS: two-stage least squares

[3.1] However, this restriction does not hold the PERR and SCCS methods.[Chapters 5.1 and 5.2] Again, the PERR and SCCS methods generally provide the ACE in a population whereas the IV analysis may not always provide the ACE, and thus it is important to consider the patients to whom the treatment effect is generalizable. In addition, it should be stressed that one of the weaknesses of the IV analysis or the PERR method is that they tend to display estimates with large standard errors relative to the conventional estimates.

In summary, there are several fundamental differences between the three chosen methods (Table 1). All three methods assume several strong assumptions to estimate exposure effects and have several pros and cons in empirical settings. The performance and application of the methods strongly depends on the nature of confounding, exposure, type of outcome, size of the sample in a particular clinical example. Therefore, correct specification and application of the methods should be carefully assessed and reported.

CONCLUSIONS

Unmeasured confounding is one of the principal problems in pharmacoepidemiological studies. In this thesis, we evaluated the performance of three statistical methods to control for unmeasured confounding: instrumental variable analysis, prior event rate ratio adjustment, and self-controlled case series design (with a focus on instrumental variable analysis). From a methodological perspective, the instrumental variable analysis is a powerful statistical tool to control for unmeasured confounding. Yet, its validity and applicability in observational studies of the effects of (pharmacological) treatments still have to be established particularly for time-varying exposure and confounders. Additionally, the estimates from instrumental variable analyses are generally less precise than those of conventional analyses, and thus the exposure effects from the instrumental variable analyses should be interpreted cautiously. The prior event rate ratio adjustment method can be applied to reduce bias due to time-fixed (measured and unmeasured) confounding. However, this method is not suitable for all possible clinical outcomes, or when confounding is time-varying. The self-controlled case-series design can also be applied to control for time-fixed (measured and unmeasured) confounding in studies of time-varying exposures and acute events. Yet, incorrect definitions of study periods may seriously bias exposure effect estimates. Therefore, we stress the importance of a complete understanding of the methods before applying them and a routine evaluation of the underlying assumptions. As the performance of the methods may differ across studies and across databases, we also stress the importance of using both statistical evidence and substantial clinical knowledge for interpretation of the study results.

REFERENCES

1. Grobbee DE, Hoes AW. Clinical epidemiology: principles, methods, and applications for clinical research. Jones & Bartlett Learning; 2009.
2. Vandenbroucke JP. When are observational studies as credible as randomised trials? *The Lancet* 2004; 363(9422): 1728-31.
3. Black N. Why we need observational studies to evaluate the effectiveness of health care. *Br Med J* 1996; 312(7040): 1215-8.
4. Brookhart MA, Stürmer T, Glynn RJ, Rassen J, Schneeweiss S. Confounding control in healthcare database research: Challenges and potential approaches. *Med Care* 2010; 48(6 SUPPL.): S114-20.
5. Baiocchi M, Cheng J, Small DS. Instrumental variable methods for causal inference. *Stat Med* 2014; 15(33(13)): 2297-340. doi:10.1002/sim.6128.
6. McMahon AD. Approaches to combat with confounding by indication in observational studies of intended drug effects. *Pharmacoepidemiol Drug Saf* 2003; 12(7): 551-8.
7. Alemayehu D, Alvir JM, Jones B, Willke RJ. Statistical issues with the analysis of nonrandomized studies in comparative effectiveness research. *Journal of managed care pharmacy : JMCP* 2011; 17(9 Suppl A): S22-26.
8. Groenwold RHH, Hak E, Hoes AW. Quantitative assessment of unobserved confounding is mandatory in nonrandomized intervention studies. *J Clin Epidemiol* 2009; 62(1): 22-8.
9. Bosco JLF, Silliman RA, Thwin SS, et al. A most stubborn bias: no adjustment method fully resolves confounding by indication in observational studies. *J Clin Epidemiol* 2010; 63(1): 64-74. doi:http://dx.doi.org/10.1016/j.jclinepi.2009.03.001.
10. Klungel OH, Martens EP, Psaty BM, et al. Methods to assess intended effects of drug treatment in observational studies are reviewed. *J Clin Epidemiol* 2004; 57(12): 1223-31.
11. Velentgas P, Dreyer NA, Nourjah P, Smith SR, Torchia MM, eds. Developing a Protocol for Observational Comparative Effectiveness Research: A User's Guide. AHRQ Publication No. 12(13)-EHC099. Rockville, MD: Agency for Healthcare Research and Quality; January 2013. www.effectivehealthcare.ahrq.gov/Methods-OCER.cfm.
12. Hernán MA, Hernández-Diaz S, Werler MM, Mitchell AA. Causal knowledge as a prerequisite for confounding evaluation: An application to birth defects epidemiology. *Am J Epidemiol* 2002; 155(2): 176-84.
13. Robins JM, Hernán MÁ, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology* 2000; 11(5): 550-60.
14. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983; 70(1): 41-55.
15. Schneeweiss S. Sensitivity analysis and external adjustment for unmeasured confounders in epidemiologic database studies of therapeutics. *Pharmacoepidemiol Drug Saf* 2006; 15(5): 291-303.
16. Schneeweiss S, Seeger JD, Smith SR. Methods for developing and analyzing clinically rich data for patient-centered outcomes research: An overview. *Pharmacoepidemiol Drug Saf* 2012; 21(SUPPL.2): 1-5.
17. Groenwold RHH, Hoes A, Nichol KL, Hak E. Quantifying the potential role of unmeasured confounders: The example of influenza vaccination. *Int J Epidemiol* 2008; 37(6): 1422-9.
18. 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; 17(3): 268-75.

19. Tannen RL, Weiner MG, Xie D. Replicated studies of two randomized trials of angiotensin-converting enzyme inhibitors: further empiric validation of the 'prior event rate ratio' to adjust for unmeasured confounding by indication. *Pharmacoepidemiol Drug Saf* 2008; 17(7): 671-85.
20. Schneeweiss S, Stürmer T, Maclure M. Case-crossover and case-time-control designs as alternatives in pharmacoepidemiologic research. *Pharmacoepidemiol Drug Saf* 1997; 6(SUPPL. 3): S51-9.
21. Maclure M, Fireman B, Nelson JC, Hua W, Shoaibi A, Paredes A, Madigan D. When should case-only designs be used for safety monitoring of medical products? *Pharmacoepidemiol Drug Saf* 2012; 21(SUPPL. 1): 50-61.
22. Farrington CP. Relative incidence estimation from case series for vaccine safety evaluation. *Biometrics* 1995; 51(1): 228-35.
23. Suissa S. The case-time-control design. *Epidemiology* 1995; 6(3): 248-53.
24. Chen Y, Briesacher BA. Use of instrumental variable in prescription drug research with observational data: a systematic review. *J Clin Epidemiol* 2011; 64(6): 687-700.
25. Rassen JA, Schneeweiss S, Glynn RJ, Mittleman MA, Brookhart MA. Instrumental variable analysis for estimation of treatment effects with dichotomous outcomes. *Am J Epidemiol* 2009; 169(3): 273-84.
26. Davies NM, Smith GD, Windmeijer F, Martin RM. Issues in the reporting and conduct of instrumental variable studies: A systematic review. *Epidemiology* 2013; 24(3): 363-9.
27. Martens EP, Pestman WR, De Boer A, Belitser SV, Klungel OH. Instrumental variables: Application and limitations. *Epidemiology* 2006; 17(3): 260-7.
28. Greenland S. An introduction to instrumental variables for epidemiologists. *Int J Epidemiol* 2000; 29(4): 722-9.
29. Hernán MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? *Epidemiology* 2006; 17(4): 360-72.
30. 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; 54(11): 3390-8.
31. Pratt N, Roughead EE, Ryan P, Salter A. Antipsychotics and the risk of death in the elderly: An instrumental variable analysis using two preference based instruments. *Pharmacoepidemiol Drug Saf* 2010; 19(7): 699-707.
32. Huybrechts KF, Gerhard T, Franklin JM, Levin R, Crystal S, Schneeweiss S. Instrumental variable applications using nursing home prescribing preferences in comparative effectiveness research. *Pharmacoepidemiol Drug Saf* 2014; 23(8): 830-8. doi:10.1002/pds.3611.
33. Davies NM, Smith GD, Windmeijer F, Martin RM. COX-2 selective nonsteroidal anti-inflammatory drugs and risk of gastrointestinal tract complications and myocardial infarction: An instrumental variable analysis. *Epidemiology* 2013; 24(3): 352-62.
34. Groenwold RHH, Hak E, Klungel OH, Hoes AW. Instrumental Variables in Influenza Vaccination Studies: Mission Impossible?! *Value Health* 2010; 13(1): 132-7.
35. Stock JH, Wright JH, Yogo M. A survey of weak instruments and weak identification in generalized method of moments. *J Bus Econ Stat* 2002; 20(4): 518-29.
36. Ionescu-Ittu R, Abrahamowicz M, Pilote L. Treatment effect estimates varied depending on the definition of the provider prescribing preference-based instrumental variables. *J Clin Epidemiol* 2012; 65(2): 155-62.
37. 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. *Can Med Assoc J* 2007; 176(5): 627.

38. Brookhart MA, Rassen JA, Schneeweiss S. Instrumental variable methods in comparative safety and effectiveness research. *Pharmacoepidemiol Drug Saf* 2010; 19(6): 537-44.
39. Boef AG, van Paassen J, Arbous MS, Middelkoop A, Vandenbroucke JP, le Cessie S, Dekkersa OM. Physician's Preference-based Instrumental Variable Analysis- Is It Valid and Useful in a Moderate-sized Study? *Epidemiology* 2014; 10.1097/EDE.0000000000000151: 00.
40. Angrist JD, Pischke JS. Mostly harmless econometrics: An empiricist's companion. Princeton Univ Pr; 2008.
41. Staiger D, Stock JH. Instrumental variables regression with weak instruments. *Econometrica* 1997; 65(3): 557-86.
42. Zivot E, Startz R, Nelson CR. Valid confidence intervals and inference in the presence of weak instruments. *International Economic Review* 1998; 39(4): 1119-44.
43. Bound J, Jaeger DA, Baker RM. Problems with instrumental variables estimation when the correlation between the instruments and the endogeneous explanatory variable is weak. *J Am Stat Assoc* 1995; 90(430): 443-50.
44. Angrist JD, Krueger AB. Instrumental variables and the search for identification: From supply and demand to natural experiments. *J Econ Perspect* 2001; 15(4): 69-85.
45. Chao J, Swanson NR. Alternative approximations of the bias and MSE of the IV estimator under weak identification with an application to bias correction. *J Econ* 2007; 137(2): 515-55.
46. Stock JH, Yogo M. Testing for weak instruments in linear IV regression. NBER Technical Working Paper No. 284, Issued in November 2002, http://www.nber.org/papers/t0284.pdf (date of access 23 July 2014).
47. Ionescu-Ittu R, Delaney JAC, Abrahamowicz M. Bias–variance trade-off in pharmacoepidemiological studies using physician-preference-based instrumental variables: a simulation study. *Pharmacoepidemiol Drug Saf* 2009; 18(7): 562-71.
48. Boef AG, Dekkers OM, Vandenbroucke JP, le Cessie S. Sample size importantly limits the usefulness of instrumental variable methods, depending on instrument strength and level of confounding. *J Clin Epidemiol* 2014; DOI:10.1016/j.jclinepi.2014.05.019.
49. Swanson SA, Hernán MA. Commentary: How to report instrumental variable analyses (suggestions welcome). *Epidemiology* 2013; 24(3): 370-4.
50. Didelez V, Meng S, Sheehan NA. Assumptions of IV methods for observational epidemiology. *Stat Sci* 2010; 25(1): 22-40.
51. Glymour MM, Tchetgen EJT, Robins JM. Credible mendelian randomization studies: Approaches for evaluating the instrumental variable assumptions. *Am J Epidemiol* 2012; 175(4): 332-9.
52. Brookhart MA, Schneeweiss S. Preference-based instrumental variable methods for the estimation of treatment effects: Assessing validity and interpreting results. *Int J Biostat* 2007; 3(1): 14.
53. Rassen JA, Brookhart MA, Glynn RJ, Mittleman MA, Schneeweiss S. Instrumental variables II: instrumental variable application-in 25 variations, the physician prescribing preference generally was strong and reduced covariate imbalance. *J Clin Epidemiol* 2009; 62(12): 1233-41.
54. Zhang Y. Cost-saving effects of olanzapine as long-term treatment for bipolar disorder. *J Ment Health Policy Econ* 2008; 11(3): 135-46.
55. Schneeweiss S, Seeger JD, Landon J, Walker AM. Aprotinin during coronary-artery bypass grafting and risk of death. *N Engl J Med* 2008; 358(8): 771-83.
56. Austin PC. A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003. *Stat Med* 2008; 27(12): 2037-49.
57. Belitser SV, Martens EP, Pestman WR, Groenwold RH, de Boer A, Klungel OH. Measuring balance and model selection in propensity score methods. *Pharmacoepidemiol Drug Saf* 2011; 20(11): 1115-29.

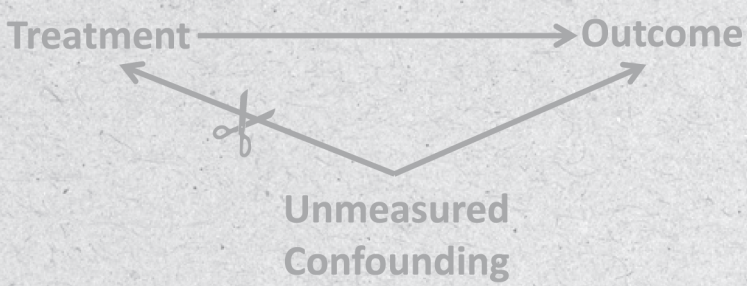
58. Davies NM, Gunnell D, Thomas KH, Metcalfe C, Windmeijer F, Martin RM. Physicians' prescribing preferences were a potential instrument for patients' actual prescriptions of antidepressants. *J Clin Epidemiol* 2013; 66(12): 1386-96.
59. Suissa S, Assimes T, Brassard P, Ernst P. Inhaled corticosteroid use in asthma and the prevention of myocardial infarction. *Am J Med* 2003; 115(5): 377-81.
60. Huiart L, Ernst P, Ranouil X, Suissa S. Low-dose inhaled corticosteroids and the risk of acute myocardial infarction in COPD. *European Respiratory Journal* 2005; 25(4): 634-9.
61. Imbens GW, Angrist JD. Identification and estimation of local average treatment effects. *Econometrica* 1994; : 467-75.
62. Landrum MB, Ayanian JZ. Causal effect of ambulatory specialty care on mortality following myocardial infarction: A comparison of propensity score and instrumental variable analyses. *Health Serv Outcomes Res* 2001; 2(3-4): 221-45.
63. Greene WH, Zhang C. Econometric analysis. 5th ed. Prentice hall New Jersey; 2003.
64. Kennedy P. A guide to econometrics. 5th ed. Cambridge, MA: MIT Press; 2003.
65. Johnston KM, Gustafson P, Levy AR, Grootendorst P. Use of instrumental variables in the analysis of generalized linear models in the presence of unmeasured confounding with applications to epidemiological research. *Stat Med* 2008; 27(9): 1539-56.
66. Foster EM. Instrumental variables for logistic regression: An illustration. *Soc Sci Res* 1997; 26(4): 487-504.
67. Terza JV, Bradford WD, Dismuke CE. The use of linear instrumental variables methods in health services research and health economics: A cautionary note. *Health Serv Res* 2008; 43(3): 1102-20.
68. Wooldridge JM. Econometric analysis of cross section and panel data. Cambridge, MA: MIT Press; 2002.
69. Palmer TM, Sterne JAC, Harbord RM, et al. Instrumental variable estimation of causal risk ratios and causal odds ratios in mendelian randomization analyses. *Am J Epidemiol* 2011; 173(12): 1392-403.
70. Schmoor C, Caputo A, Schumacher M. Evidence from nonrandomized studies: A case study on the estimation of causal effects. *Am J Epidemiol* 2008; 167(9): 1120-9.
71. Ramirez SPB, Albert JM, Blayney MJ, et al. Rosiglitazone is associated with mortality in chronic hemodialysis patients. *Journal of the American Society of Nephrology* 2009; 20(5): 1094-101.
72. Bosco JLF, Silliman RA, Thwin SS, et al. A most stubborn bias: no adjustment method fully resolves confounding by indication in observational studies. *J Clin Epidemiol* 2010; 63(1): 64-74.
73. Burgess S. Statistical issues in Mendelian randomization: use of genetic instrumental variables for assessing causal associations. 2011.
74. Rassen JA, Brookhart MA, Glynn RJ, Mittleman MA, Schneeweiss S. Instrumental variables I: instrumental variables exploit natural variation in nonexperimental data to estimate causal relationships. *J Clin Epidemiol* 2009; 62(12): 1226-32.
75. Garabedian LF, Chu P, Toh S, Zaslavsky AM, Soumerai SB. Potential Bias of Instrumental Variable Analyses for Observational Comparative Effectiveness Research. *Ann Intern Med* 2014; 161(2): 131-8.
76. Uddin MJ, Groenwold RHH, de Boer A, Belitser S, Roes KCB, Hoes AW, Klungel OH. Performance of instrumental variable methods in cohort and nested case–control studies: a simulation study. *Pharmacoepidemiol Drug Saf* 2014; 23: 165–177.
77. Yu M, Xie D, Wang X, Weiner MG, Tannen RL. Prior event rate ratio adjustment: Numerical studies of a statistical method to address unrecognized confounding in observational studies. *Pharmacoepidemiol Drug Saf* 2012; 21(SUPPL.2): 60-8.
78. Tannen RL, Weiner MG, Xie D. Use of primary care electronic medical record database in drug efficacy research on cardiovascular outcomes: Comparison of database and randomised controlled trial findings. *BMJ (Online)* 2009; 338(7691): 395-9.

79. Tannen R, Xie D, Wang X, Yu M, Weiner MG. A new "Comparative Effectiveness" assessment strategy using the THIN database: Comparison of the cardiac complications of pioglitazone and rosiglitazone. *Pharmacoepidemiol Drug Saf* 2013; 22(1): 86-97.
80. Weiner MG, Xie D, Tannen RL. Replication of the Scandinavian Simvastatin Survival Study using a primary care medical record database prompted exploration of a new method to address unmeasured confounding. *Pharmacoepidemiol Drug Saf* 2008; 17(7): 661-70.
81. Whitaker HJ, Farrington CP, Spiessens B, Musonda P. Tutorial in biostatistics: The self-controlled case series method. *Stat Med* 2006; 25(10): 1768-97.
82. Whitaker HJ, Hocine MN, Farrington CP. The methodology of self-controlled case series studies. *Stat Methods Med Res* 2009; 18(1): 7-26. doi:10.1177/0962280208092342 [doi].
83. Weldeselassie YG, Whitaker HJ, Farrington CP. Use of the self-controlled case-series method in vaccine safety studies: Review and recommendations for best practice. *Epidemiol Infect* 2011; 139(12): 1805-17.
84. Whitaker H. The self controlled case series method. *BMJ* 2008; 337:a1069.
85. Douglas IJ, Langham J, Bhaskaran K, Brauer R, Smeeth L. Orlistat and the risk of acute liver injury: Self controlled case series study in UK Clinical Practice Research Datalink. *BMJ (Online)* 2013; 346:f1936(7906).
86. Hubbard R, Farrington P, Smith C, Smeeth L, Tattersfield A. Exposure to tricyclic and selective serotonin reuptake inhibitor antidepressants and the risk of hip fracture. *Am J Epidemiol* 2003; 158(1): 77-84.
87. Langan SM, Minassian C, Smeeth L, Thomas SL. Risk of Stroke Following Herpes Zoster: A Self-Controlled Case-Series Study. *Clinical Infectious Diseases* 2014; : ciu098.
88. Wijlaars LPMM, Nazareth I, Whitaker HJ, Evans SJW, Petersen I. Suicide-related events in young people following prescription of SSRIs and other antidepressants: A self-controlled case series analysis. *BMJ Open* 2013; 3:e003247(9). doi:10.1136/bmjopen-2013-003247.
89. Ramsay EN, Pratt NL, Ryan P, Roughead EE. Proton pump inhibitors and the risk of pneumonia: A comparison of cohort and self-controlled case series designs. *BMC Medical Research Methodology* 2013; 13:82,doi:10.1186/1471-2288-13-82(1). doi:10.1186/1471-2288-13-82.
90. Paddy Farrington C, Anaya-Izquierdo K, Whitaker HJ, Hocine MN, Douglas I, Smeeth L. Self-controlled case series analysis with event-dependent observation periods. *Journal of the American Statistical Association* 2011; 106(494): 417-26.
91. Hua W, Sun G, Dodd CN, et al. A simulation study to compare three self-controlled case series approaches: Correction for violation of assumption and evaluation of bias. *Pharmacoepidemiol Drug Saf* 2013; 22(8): 819-25.
92. Greenland S, Robins JM, Pearl J. Confounding and collapsibility in causal inference. *Stat Sci* 1999; : 29-46.
93. Burgess S, Granell R, Palmer TM, Sterne JAC, Didelez V. Lack of identification in semiparametric instrumental variable models with binary outcomes. *Am J Epidemiol* 2014; 180(1): 111-9.
94. Robins JM. Correcting for non-compliance in randomized trials using structural nested mean models. *Commun Stat* 1994; 23(8): 2379-412.
95. Austin PC. A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003. *Stat Med* 2008; 27(12): 2037-49.
96. Greenland S. Basic methods for sensitivity analysis of biases. *Int J Epidemiol* 1996; 25(6): 1107-16.
97. Sturmer T, Schneeweiss S, Rothman KJ, Avorn J, Glynn RJ. Performance of propensity score calibration--a simulation study. *Am J Epidemiol* 2007; 165(10): 1110-8. doi:kwm074 [pii].
98. Sturmer T, Schneeweiss S, Avorn J, Glynn RJ. Adjusting effect estimates for unmeasured confounding with validation data using propensity score calibration. *Am J Epidemiol* 2005; 162(3): 279-89. doi:kwi192 [pii].

99. Miettinen OS. The need for randomization in the study of intended effects. *Stat Med* 1983; 2(2): 267-71.
100. Joseph K, Mehrabadi A, Lisonkova S. Confounding by Indication and Related Concepts. *Current Epidemiology Reports* 2014; 1(1): 1-8.
101. Musonda P, Hocine MN, Whitaker HJ, Farrington C. Self-controlled case series analyses: Small-sample performance. *Comput Stat Data Anal* 2008; 52(4): 1942-57.
102. Fang G, Brooks JM, Chrischilles EA. Apples and oranges? Interpretations of risk adjustment and instrumental variable estimates of intended treatment effects using observational data. *Am J Epidemiol* 2011; 175(1): 60-5. doi:10.1093/aje/kwr283.

Unmeasured confounding IV PERR SCCS Observational studies
PROTECT Causal Inference Statistical methods পরিসংখ্যানগত পদ্ধতি
Statistische methoden IV PERR SCCS Pseudo Randomization
Pharmacoepidemiology IV Analyses PERR SCCS
General Practice Databases Residual Confounding
Electronic Healthcare Databases Epidemiology
Time-fixed Confounding Time-fixed Exposure
Time-varying Exposure
Time-varying Confounding
Non-randomized Studies
IV PERR SCCS
PROTECT 2SLS
IV IV

APPENDICES



APPENDIX A

Summary

Randomized controlled trials are not always feasible to estimate the effect of pharmacological treatments, for instance, to detect rare adverse drug reactions. In that case, observational (non-randomized) studies are essential. However, in observational studies, the treatment effect can be biased due to possible confounding by relevant patients' characteristics that are not equally distributed over the treatment groups. Several statistical methods, e.g., multivariable regression, have been used to control for confounding by measured confounders. Nevertheless, confounding by unmeasured patient characteristics (e.g., disease severity) is also a potential problem of observational studies. Alternative methods have therefore been proposed to control for unmeasured confounding. In this thesis, we focus on three of those methods: instrumental variable (IV) analysis, prior event rate ratio (PERR) adjustment, and the self-controlled case series (SCCS) design.

Chapter I provides a general introduction to the thesis. In Chapter II, the conceptual framework of IV analysis and why application of IV analysis in pharmacoepidemiology may be hampered are discussed (Chapter 2.1). We also reviewed and described the methods (e.g., two-stage least squares, generalized method of moments) for IV estimation for continuous as well as binary outcomes, exposures, and IVs and provide guidance on when different methods may be applicable together, each of which have advantages and disadvantages (Chapter 2.2).

Chapter III contains two simulation studies of IV analysis. The first (Chapter 3.1) focuses on the performance of IV estimates for different combinations of continuous or binary IV, exposure, and outcome in the cohort as well as the nested case-control (NCC) designs. For all types of IV and exposure, IV estimates are very unstable and biased when the IV is very weakly associated with the exposure. In addition, the selection and construction of IVs in a NCC study are in line with a cohort study. However, IV estimates in the NCC study are more variable than those of a cohort study, which can be partly remedied by increasing the number of cases, or by increasing the number of controls per case. In the second simulation study (Chapter 3.2), the performance of balance measures commonly used in propensity score methods (i.e., standardized difference) was assessed to falsify the third assumption of IV analysis. The third assumption states that the IV is independent of confounders. This study showed that the bias of IV estimates increased with increasing values of the standardized difference (i.e., decreasing balance of confounders across IV levels). Hence, it is argued that the standardized difference can be a useful tool to falsify this assumption though the balance of measured confounders between IV categories may not guarantee the balance of unmeasured confounders.

In Chapter IV, two empirical studies of IV analysis in which the exposures of interest are either time-fixed or time-varying are discussed. In the first study (Chapter 4.1), different physician's prescribing preference based IVs are evaluated in two databases (CPRD, UK and Mondriaan, Netherlands) in a study of inhaled long-acting beta2-agonist use and the risk of myocardial infarction. In the second study (Chapter 4.2), the performance of IV analysis is assessed using different IVs in multiple databases (THIN, UK; BIFAP, Spain; and Mondriaan,

Netherlands) in a study of antidepressant use and the risk of hip fracture. In both empirical studies, we did not identify any IV that is consistently valid across the databases for time-fixed exposures. Moreover, none of the IVs we considered appeared to be valid for time-varying exposures in all databases. We observed that the performance of IV analyses varied across the databases and strongly depends on the definition of IVs and size of the samples. Therefore, a clear strategy to evaluate the (apparent) validity of possible IVs should be included in the evaluation process of every IV analysis.

In Chapter V, the performance of the PERR adjustment method is assessed in a simulation study. Moreover, the impact of violations of the assumptions of the SCCS design and different definitions of observation periods in SCCS analyses are studied in an empirical study of antidepressants use and risk of hip fracture using two general practice databases (Mondriaan, Netherlands and THIN, UK). The PERR method can be applied to reduce bias due to unmeasured confounding; however, in particular situations (e.g., when confounders change between prior and post periods, prior events influence the probability of exposure) it may produce biased exposure effects (Chapter 5.1). The SCCS analyses show that the incorrect definitions of observation periods (e.g., starting follow-up with the first prescription during the study period, or when subjects are censored at the event) may seriously bias exposure effect estimates (Chapter 5.2).

In Chapter VI, the results of the different studies mentioned above are combined with the existing literature to provide recommendations for the application of the methods to control for unmeasured confounding in pharmacoepidemiology. Although IV analysis appears a powerful statistical tool to control for unmeasured confounding, its validity and applicability in observational studies of the effects of (pharmacological) treatments still have to be established particularly for studies of time-varying exposures. The PERR method and the SCCS design can be applied for time-varying exposure to reduce bias due to time-fixed measured and unmeasured confounding. However, they may provide biased results when confounders change over time or they may not be suitable for all possible clinical outcomes. Therefore, we emphasize the routinely evaluation of the assumptions of these methods. As the performance of the methods may differ across studies and across databases, we also stress the importance of using both statistical evidence and substantial clinical knowledge for interpretation of the study results.

APPENDIX B

Samenvatting

Het is niet altijd mogelijk om met behulp van gerandomiseerd onderzoek de effecten van farmacologische behandelingen te schatten, bijvoorbeeld als het gaat om het vaststellen van zeldzame bijwerkingen. In dat geval zijn observationele (niet-gerandomiseerde) onderzoeken essentieel. Echter, in observationeel onderzoek kan het waargenomen effect van de behandeling worden beïnvloed door verstorende variabelen (bv prognostische kenmerken van patiënten) die niet gelijk verdeeld zijn over de behandelingsgroepen (confounding). Verschillende statistische methoden, bijvoorbeeld multivariabele regressie analyse, kunnen worden gebruikt om te corrigeren voor confounding door de gemeten confounders. Toch kunnen *ongemeten* patiëntkarakteristieken (bijvoorbeeld ernst van ziekte) dan nog steeds leiden tot confounding in observationeel onderzoek. Daarom zijn alternatieve methoden voorgesteld om te controleren voor ongemeten confounders. Dit proefschrift richt zich op drie van deze methoden: instrumentele variabele (IV)-analyse, correctie op basis van de *prior event rate ratio* (PERR), en het *self-controlled case series* (SCCS) design.

Hoofdstuk I geeft een algemene inleiding op het proefschrift. In hoofdstuk II, wordt het concept van IV-analyse besproken alsmede mogelijke belemmeringen in de toepassing van IV-analyse in farmaco-epidemiologisch onderzoek (hoofdstuk 2.1). Daarnaast wordt een overzicht gegeven van verschillende methoden voor IV-analyse (bijvoorbeeld *two-stage least squares* en *generalized method of moments*) voor continue en binaire uitkomsten, blootstellingen en instrumentele variabelen (hoofdstuk 2.2). De voor- en nadelen van de verschillende methoden worden besproken.

Hoofdstuk III beschrijft twee simulatiestudies die zich richten op IV-analyse. De eerste (hoofdstuk 3.1) is gericht op de validiteit en precisie van IV schattingen voor verschillende combinaties van continue of binaire IVs, blootstellingen en uitkomsten in zowel cohort als geneste patiënt-controle (NCC) onderzoeken. In alle gevallen zijn de IV schattingen zeer instabiel en vertekend wanneer de IV een zwakke relatie met de blootstelling heeft. De operationalisatie van IVs in een NCC onderzoek is hetzelfde als in een cohort onderzoek. Echter, schattingen van de relatie tussen blootstelling en uitkomst op basis van IV-analyse zijn in een NCC onderzoek meer variabel dan in een cohort onderzoek. Dit kan deels worden verholpen door het vergroten van het aantal patiënten, en/of door het vergroten van het aantal controles per patiënt. In de tweede simulatiestudie (hoofdstuk 3.2) wordt een balansmaat die veel wordt gebruikt in propensity score methoden (namelijk het gestandaardiseerde verschil) beoordeeld op zijn mogelijkheid om de derde aanname van IV-analyse te evalueren. Deze aanname is dat de IV onafhankelijk is van confounders. Deze studie toonde aan dat de *bias* van IV schattingen toeneemt met toenemende waarden van het gestandaardiseerde verschil (dat wil zeggen, met een verminderde balans van confounders over verschillende IV niveaus). De conclusie is daarom dat het gestandaardiseerde verschil een nuttig instrument kan zijn om de derde aanname van IV-analyse te evalueren.

In hoofdstuk IV worden twee empirische onderzoeken beschreven waarin IV-analyse wordt toegepast en de blootstellingen stabiel zijn of variëren over de tijd. In het eerste

onderzoek (hoofdstuk 4.1) worden verschillende IVs geëvalueerd die zijn gebaseerd op het voorschrijfgedrag van artsen. Het betreft een onderzoek in twee databases (CPRD (UK) en Mondriaan (Nederland)) naar de associatie tussen het gebruik van geïnhaleerde langwerkende bèta-2-agonisten en het risico op een myocardinfarct. In het tweede onderzoek (hoofdstuk 4.2) worden de validiteit en toepasbaarheid van IV-analyse beoordeeld in verschillende databases (THIN (UK), BIFAP (Spanje) en Mondriaan (Nederland)) in een onderzoek naar het gebruik van antidepressiva en het risico op een heupfractuur. In beide onderzoeken werden geen IVs gevonden die universeel geschikt lijken voor het bestuderen van blootstellingen die niet variëren in de tijd, noch voor blootstellingen die wel variëren in de tijd. De prestaties van IV-analyses variëren tussen databases en zijn sterk afhankelijk van de definitie van de IV en de omvang van de datasets. Daarom is het van belang om voorafgaand aan een analyse een duidelijke strategie te beschrijven om de (schijnbare) validiteit van IVs vast te stellen.

In hoofdstuk V wordt de prestatie van de PERR correctiemethode onderzocht in een simulatiestudie. De PERR methode kan worden toegepast om te corrigeren voor ongemeten confounders, echter in bepaalde situaties (bijvoorbeeld wanneer confounding varieert in de tijd, of wanneer het optreden van klinische eindpunten voorafgaand aan een periode van blootstelling invloed heeft op de kans op blootstelling) kan de methode tot bias leiden (hoofdstuk 5.1). Daarnaast zijn de verschillende uitgangspunten van het SCCS design onderzocht in een empirisch onderzoek naar het gebruik van antidepressiva en het risico op een heupfractuur met twee databases (THIN (UK) en Mondriaan (Nederland)) (hoofdstuk 5.2). Deze analyses laten zien dat in dit design een onjuiste definitie van de observatieperiode (bijvoorbeeld als voorschrijven van een geneesmiddel de start van follow-up definieert, of als de observatietijd na een klinische eindpunt wordt genegeerd) kan leiden tot ernstige bias van de schattingen van de effecten van de blootstelling.

In hoofdstuk VI worden de resultaten van de verschillende hierboven beschreven onderzoeken gecombineerd met de bestaande literatuur om te komen tot aanbevelingen ten aanzien van de drie methoden om te controleren voor ongemeten confounding in farmaco-epidemiologisch onderzoek. Hoewel IV-analyse potentieel een krachtig statistisch instrument is om te controleren voor ongemeten confounding moet zijn toepasbaarheid in observationele studies naar de effecten van (farmacologische) behandelingen nog grotendeels worden vastgesteld. Dit geldt in het bijzonder voor onderzoek naar tijdsafhankelijke blootstellingen. De PERR methode en het SCCS design kunnen worden toegepast in onderzoek naar tijdsafhankelijke blootstellingen. Ze kunnen echter vertekende resultaten geven wanneer confounders veranderen na verloop van tijd. Bovendien zijn ze niet voor alle mogelijke klinische eindpunten toepasbaar. Het is daarom van belang om routinematig de uitgangspunten van deze methoden te evalueren. Aangezien de prestaties van de methoden kunnen verschillen tussen verschillende onderzoeken en tussen verschillende databases, is het voor de interpretatie van de resultaten van observationele studies naar effecten van geneesmiddelen van belang om naast statistische gegevens klinische en inhoudelijke kennis van de database te gebruiken.

APPENDIX C

Acknowledgement

First and above all, I praise the almighty Allah (God), who has given me this opportunity to work with the incredibly amazing and inspirational team of supervisors: Prof. Dr. A. de Boer, Prof. Dr. Kit C.B. Roes, Dr. R.H.H. Groenwold and Dr. O.H. Klungel and has granted me to finish a new stage of my life successfully by doing this PhD research. I am highly grateful to my supervisors for allowing me to work under their guidance.

Dear Prof. Ton, my esteemed promoter, first, I express my sincere gratitude to you for accepting me as a PhD student, your warm encouragement, thoughtful guidance, valuable comments, clinical inputs, and correction of the thesis. You are a very kind person that I ever meet; you gave me a lot of opportunities to access your office and discuss my problems with you. Although you are very busy, you replied my emails quickly and provided your comments and suggestions on the manuscripts rapidly. I greatly appreciate your excellent assistance regarding all official letters. I found you not only a good supervisor and researcher but also a good administrator.

Dear Prof. Kit, my respected promoter, I want to express my deep thanks to you for accepting me as a PhD student, all of your inspirations and new ideas in my thesis during last four years. You gave me ample opportunities to discuss my works with you, which definitely helped to improve my manuscripts and statistical analyses. Without your critical statistical comments in all manuscripts, the thesis would not have been of such standard.

Dear Dr. Rolf, I owe a deep sense of gratitude to you for your untiring patience and attention given to me whilst completing this PhD thesis. I am really lucky to find you as my daily supervisor. I am grateful to you for encouraging me to push my limits, for supporting me in pursuing my ideas. You were always willing to meet weekly, discuss ideas and provide insightful thought on my works. I appreciate your excellent and extensive feedback on the earlier version of my manuscripts, which indeed assisted me to improve the quality of my thesis. I have learnt a lot about epidemiology and Pharmacoepidemiology from you. Not only you have an excellent supervising capacity, but you have also a great teaching quality. I also thank you for translating my thesis summary into Dutch.

Dear Dr. Olaf, I would like to extend my best thanks to you for selecting me under this PhD position within the PROTECT project. The day I first talked with you, I was really impressed! I also want to express my profound gratitude for your encouragement, brilliant new ideas and key motivations throughout my PhD. Your excellent feedback on the earlier version of my manuscripts gave me a tremendous opportunity to improve the quality of my thesis. Last four years, I have enjoyed working with you. I learned a lot from you how to think positively in all aspects. Finally, I am grateful to you for financial supports of my PhD project.

Dear Prof. Arno Hoes, I first met with you during the introduction of epidemiology course. I remember your excellent explanation on case-control study. Later on we met every three months in PROTECT PhD meeting. I am so honoured that you have been a co-author of almost all manuscripts of my thesis. I like to express my great thanks to you for your

contributions, critical clinical comments on the manuscripts. I have learnt from you, how to use less jargon words in the manuscript.

Dear Dr. Wiebe, I would like to thank you for your contributions and constructive feedback at the beginning of my PhD and in one of my articles. I appreciate your ideas regarding the falsification of instrumental variable assumption in the case of continuous IV.

Dear Svetlana, I would like to give my sincere thanks to you for supporting me during the past four years. Your efforts regarding statistical issues such as making R-codes, comments on the complex statistical equations, making nice plots for manuscripts are really appreciated. We have shared many ideas of social aspects, discussed family issues as well as Russian and Bangladeshi culture, which are indeed memorable to me.

To all my co-authors: Dr. Helga Gardarsdottir, Dr. Patrick C Souverein, Dr. Nicolle Gatto, Dr. Consuelo Huerta, Dr. Elisa Martin, Dr Sanni Ali, Elena Rivero-Ferrer, Dr. Ana SM Afonso, Dr. Paola Primatesta, Dr. Claudia Becker, Gianmario Candore, Dr. Mark CH de Groot, and Yolanda Alvarez, thank you for your valuable contributions in the manuscripts. Moreover, I would like to give my special thanks to Dr. Patrick C Souverein, Dr. Claudia Becker, Dr. Elisa Martin and Gianmario Candore for providing me several data sets and continuous feedback on the data analyses. In addition, I avail this opportunity to express my thanks to all databases organizations (Mondriaan, CPRD, THIN, BIFAP) for allowing me to use their data.

I would like to express my cordial thanks to Prof. Dr. H.G.M. Leufkens, Prof. Dr. T.P. van Staa, Prof. Dr. D.E. Grobbee, Prof. Dr. Jan P. Vandenbroucke, and Dr. M.J.C. Eijkemans for being members of my thesis reading committee and for the time you have spent reading my thesis.

I want to convey my cordial thanks to the members, Dr. Rolf, Floriaan, Grace, Marijn, Mirjam, and Stavros, of the bi-weekly causality meetings for sharing their experience and knowledge on different issues of causality. Special thanks are also given to Dr. Rolf for his initiative to create this journal club. Floriaan, I remember the time that we spent together during the ICPE 2013 in Montreal, particularly in Ottawa airport where we missed our flight.

I am thankful to ISPE for providing me scholarship in order to attend and present my research at the ICPE 2013, Montreal, ISPE mid-year meeting 2014, Netherlands, and ICPE 2014, Taipei.

I would like to extend my gratitude to my AIO friends, colleagues and staffs at the Division of Pharmacoepidemiology and Clinical Pharmacology for their cordial help to improve my knowledge in Pharmacoepidemiology. I have enjoyed working in this division in which an excellent research environment exists. I am thankful to my Dutch colleagues, particularly, Francisco, Rianne, Heshu, Arjo, Vikash, and Hilda for their valuable help regarding translation of several letters from Dutch language, making appointments, filling out tax return forms and so on. Moreover, I am thankful to Hans for providing me several practical information

of thesis printing and find out sponsors. In addition, I express my thanks to Arief for his help to correct SAS codes. I also extend my appreciation to my all international friends, Sanni, Yared, Hamid, Sulmaz, Yaser, Fawaz, Mohammad, Ali, Alfi, Turki, Fariba, Yaumo, Grace, Teresa for sharing with me their ideas regarding PhD research, country experiences, funny stories, foods, etc. I must admit it was a great time with you and a nice memory for me. I wish you all the best in your study. Yaser, thank you for accepting my request to be one of my paranymps. Jet and Rianne, thank you for your cooperation during the EU2P online teaching in 2013 and 2014.

Dear Ineke, Anja and Suzanne thank you so much for all of your supports last four years. I appreciate your help regarding all official letters and appointments that you have arranged for me. I express my thanks to the Monique, Haks, Bianca, Nina Matitaputty, Anneke for their helps in several official purposes. Moreover, I want to thank Ines for her supports regarding the manuscript distribution to PROTECT consortium and making meeting schedule related to my PhD project.

Dear Willem Rekvelt, I express my cordial thanks for your invaluable friendly help during my PhD. Whenever I requested you my demand related to ICT, you tried your best to fulfil it. You improved the computer performance in the data lab that really facilitated to finish my data analyses on time.

Sanni, in October 2010, you found me in UUING group when I was searching my house in Utrecht. Your help regarding my housing is appreciated. I extend my cordial thanks to you for a nice collaboration with me in last 4 years. We have also worked together in two joint projects and thank you for your contributions. We did many courses together, visited several conferences, and countries that are memorable to me. The complexity of visa issues that we faced so far were not forgettable though finally we succeeded to get the visa. I like to thank you for accepting my request to be one of my paranymps. I am grateful to you as you have shared your experiences with me related to thesis printing, applying for funding and other ceremonial things. We have shared many aspects of life, family matters, two countries culture, and so on, that are also memorable to me. I am happy that you have already started your post-doctoral research; I wish you all the best in your future life.

To my other friends in Utrecht, Mazda, Farshad, Mojtaba, Mohadese, Amr, Mohammed, it was great to share many experiences with you. I thank you all and wish you all the best in your PhD study. I am grateful to my Bangladeshi friend Kamal in Utrecht for his continuous supports in last 4 years, particularly, to help me to manage my housing. I would especially like to thank Jet and Nico (Nieuwegein), Femke, Hilda, and Mimount for their kind social help. Jet, when I was staying in your house, I was impressed with your cordial behaviour. I wish you a good health and happy time.

I would like to convey my gratitude to my employer in Bangladesh (Department of Statistics, Shahjalal University of Science and Technology, Sylhet) for providing me study leave to

finish this PhD. I would also like to thank the dean and head of my department, Prof. Dr. Zakir Hossain, Prof. Dr. Sabina Islam and Prof. Dr. Ahmad Kabir, for their support on several official letters of study leave.

I also have to thank my all well-wishers, especially, Prof. Rahmat Ali, Dr. Nazrul Islam, Prof. Taj Uddin, Dr. Ohid Ullah, Dr. Ismail Hossain, Dr. Rezaul Islam, Dr. Abu Yousuf, Faruque (PGE), Atique, Baker (USA), Rupok, Mohib, Selim, Anam, Mohi without whose encouragement the completion of this work would have been difficult. Ohid Bhai, I want to thank you for your invaluable supports when you were living in Netherlands.

I want to express my gratitude to my family members, especially my parents, parents-in-law, brothers, Dr. Iqbal and Asad, brothers-in-law, Toufiq, Delower, Masud and Saidul Haque, sisters, Nazu, Kamran and Luthfa, and sisters-in-law, Lubna, Naima, and Fariha, uncles, Khairul Bashar, Dr. Nasar Uddin Ahmed (FIU, USA), Dr. Omar Faruque, and Md. Humayn Kabir, for their continuous encouragement during the period of this study.

I am greatly indebted to my beloved wife, Jannat, for her encouragement, support and sacrifices without any complaint throughout the study. To my lovely sweet daughter Nudrat, at home you always ask me many questions, e.g., Abbu, what are you doing in the university? In fact, you make my time enjoyable and you are the origin of my happiness.

Last but not least, I wish to express my sincere thanks to all those who have one way or another helped me in making this study a success.

Dank u wel, সবাই কে অনেক অনেক ধন্যবাদ (Sobai ke onek onek dhonnobad)

APPENDIX D

List of Co-authors

Co-authors in this thesis are listed in alphabetical order as:

ANA SM AFONSO, PhD
- Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, Utrecht, The Netherlands

MOHAMMED SANNI ALI, PhD
- Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, Utrecht, The Netherlands

YOLANDA ALVAREZ
- European Medicines Agency, EMA, London, United Kingdom

CONSUELO H ALVAREZ, MD, PhD
- BIFAP Research Unit, Division of Pharmacoepidemiology and Pharmacovigilance, Agencia Española de Medicamentos y Productos Sanitarios (AEMPS), Madrid, Spain

CLAUDIA BECKER, PhD
- Basel Pharmacoepidemiology Unit, Division of Clinical Pharmacy & Epidemiology, Department of Pharmaceutical Sciences, University of Basel, Basel, Switzerland

SVETLANA V BELITSER, MSc
- Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, Utrecht, the Netherlands

ANTONIUS DE BOER, MD PhD
- Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, Utrecht, the Netherlands

ELENA RIVERO-FERRER
- Global Clinical Epidemiology, Novartis Farmaceutica S.A., Barcelona, Spain

HELGA GARDARSDOTTIR, PhD
- Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, Utrecht, the Netherlands
- Department of Clinical Pharmacy, Division of Laboratory and Pharmacy, University Medical Center Utrecht, Utrecht, Netherlands

NICOLE GATTO, PhD
- Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, USA
- Epidemiology, Worldwide Safety Strategy, Pfizer Inc, New York, USA

ROLF HH GROENWOLD, MD PhD
- Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, the Netherlands
- Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, Utrecht, the Netherlands

MARK CH de GROOT, PhD
- Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, Utrecht, the Netherlands

ARNO W HOES, MD PhD
- Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, the Netherlands

OLAF H KLUNGEL, PharmD PhD
- Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, Utrecht, the Netherlands
- Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, the Netherlands

ELISA M MERINO, PhD
- BIFAP Research Unit, Division of Pharmacoepidemiology and Pharmacovigilance, Agencia Española de Medicamentos y Productos Sanitarios (AEMPS), Madrid, Spain

WIEBE R PESTMAN
- Department of Mathematical Psychology,
  Catholic University of Leuven, Leuven, Belgium

PAOLA PRIMATESTA, PhD
- Global Epidemiology, Novartis Pharma AG, Basel, Switzerland

KIT CB ROES, PhD
- Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, the Netherlands

TP van STAA, PhD
- London School of Hygiene and Tropical Medicine, London, UK

PATRICK C SOUVEREIN, PhD
- Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, Utrecht, the Netherlands

APPENDIX E

List of Publications

Publications related to this thesis

1. **Md Jamal Uddin**, Rolf HH Groenwold, Anthonius de Boer, Svetlana V Belitser, Kit CB Roes, Arno W Hoes, Olaf H Klungel. Performance of instrumental variable methods in cohort and nested case-control studies: a simulation study. *Pharmacoepidemiology and Drug Safety, 2014, 23(2):165–177.*

2. M. Sanni Ali/**Md Jamal Uddin***, Rolf HH Groenwold, Wiebe R Pestman, Svetlana V Belitser, Arno W Hoes, Anthonius de Boer, Kit CB Roes, Olaf H Klungel. Quantitative falsification of instrumental variables assumption using balance measures. *Epidemiology, 2014, 25(5):770-772*.
   *Both authors contributed equally to this work and are listed in alphabetical order

3. Rolf HH Groenwold, **Md Jamal Uddin**, Kit CB Roes, Anthonius de Boer, Elena Rivero-Ferrer, Elisa Martin, Nicolle M Gatto, Olaf H Klungel. Instrumental variable analysis in randomized trials with non-compliance and observational pharmacoepidemiological studies. *OA Epidemiology 2014 May 09; 2(1):9.*

4. **Md Jamal Uddin**, Rolf HH Groenwold, Tjeerd P. van Staa, Anthonius de Boer, Svetlana V Belitser, Arno W Hoes, Kit CB Roes, Olaf H Klungel. Performance of Prior Event Rate Ratio Adjustment Method in Pharmacoepidemiology: A Simulation Study, *Pharmacoepidemiology and Drug Safety, 2014, DOI: 10.1002/pds.3724.*

Publications unrelated to this thesis

1. **Md Jamal Uddin**, Md Zakir Hossain. Predictors of infant mortality in a developing country. *Asian Journal of Epidemiology, 2008; 1(1): 1-16.*

2. Mohammed Taj Uddin, Md Nazrul Islam, **Md Jamal Uddin**. A survey on nutrition knowledge of physicians in Bangladesh: evidence from Sylhet data. *South East Asian Journal of Medical Education. 2008; 2(2):14-17.*

3. Syed Md Fakrul ahsan, Md Nazrul Islam, **Md Jamal Uddin**, Mohammed Taj Uddin. Statistical modeling of groundwater arsenic contamination level of Bangladesh with chemical elements. *Journal of Applied Quantitative Methods, 2008; 3(3): 254-262.*

4. Mohammed Morad, Mohammed Nasir Uddin, **Md Jamal Uddin**, Mohammed Mostafa Kamal. Involvement of women in working place and its impact on fertility: a study in Sylhet city. *SIU Studies, 2008; 1(3).*

5. Mohammad Ohid Ullah, **Md Jamal Uddin**. A Study on Evolution of the pH Level over Time in Patients Suffering from Reflux. *Shiraz E Medical Journal, 2008; 9(3):141-148.*

6. **Md Jamal Uddin**, Md Zakir Hossain, Mohammad Ohid Ullah. Child mortality in a developing country-a statistical analysis. *Journal of Applied Quantitative Methods, 2009; 4 (3): 270-283*.

7. Mohammad Ohid Ullah, **Md Jamal Uddin**, Mohammad M Rahman, Md Nazrul Islam, Mohammed Taj Uddin. A Study to detect the seasonal effect of chickenpox in Bangladesh. *Romanian Statistical Review (medical statistics), 2009; nr. 12*.

8. Halima Jahan, Osul Ahmed Chowdhury, **Md Jamal Uddin**. Study of seroepidemiology of HEV and its association in new entrants and final year medical students-a study on Sylhet MAG Osmani Medical College, Bangladesh. *Proceedings of the Pakistan Academy of Sciences, 2009; 46(2): 69-74*.

9. Halima Jahan, Osul Ahmed Chowdhury, **Md Jamal Uddin**. Study of seroepidemiology of H. pylori infection and their association in new entrants and final year students of Sylhet MAG Osmani Medical College, Bangladesh. *International Journal of Medicine and Medical Sciences, 2010; 2(11): 354-358*.

10. Muhammad AB Chowdhury, Mohammed Taj Uddin, **Md Jamal Uddin**. Oil seeds area and production variability in Bangladesh. *Journal of Applied Quantitative Methods, 2014; 9(2): 50-57*.

Conference abstracts related to this thesis*

1. Instrumental variables analysis using multiple databases: an example of antidepressant use and risk of hip/femur fracture. Oral presentation at 30th International Conference on Pharmacoepidemiology & Therapeutic Risk Management (ICPE), 2014, Taiwan and Netherlands Epidemiological Society (WEON), 2014.

2. Evaluating different physician's prescribing preference based instrumental variables in the study of beta2-agonist use and the risk of acute myocardial infarction. Oral presentation at the 30th ICPE 2014, Taiwan and poster presentation at the WEON 2014, Netherlands.

3. The use of prior event rate ratio adjustment method for controlling unmeasured confounding in pharmacoepidemiologic studies: a cautionary note (selected as best method abstract). Oral presentation at the 30th ICPE, 2014, Taiwan.

4. Self-controlled case series design: an example of antidepressant use and the risk of hip fracture. Oral presentation at the WEON 2014 and poster presentation at the 30th ICPE, 2014, Taiwan.

5. Performance of prior event rate ratio adjustment method in pharmacoepidemiology. Oral presentation at the international society for pharmacoepidemiology (ISPE) Mid-Year Meeting, 2014, Netherlands and poster presentation at the WEON 2014, Netherlands.

6. Quantitative verification of instrumental variables assumption using balance measures. Poster presentation at the Society of Epidemiologic Research (SER), 2013, USA, 29th ICPE 2013, Canada, and WEON 2013, Netherlands.

7. Application of instrumental variables analysis: an example of beta2-agonist use and myocardial infarction. Oral presentation at the 29th ICPE 2013, Canada and poster presentation at the WEON 2013, Netherlands.

8. Performance of instrumental variable methods in case-control and cohort studies: a simulation study. Oral presentation at the WEON, 2012 and poster presentation at the 28th ICPE, 2012, Spain.
   **Authors names of these abstracts have been listed in the different chapters of this thesis*

APPENDIX F

About the author

Md Jamal Uddin was born in Comilla, Bangladesh, on the 1st of August 1977. From November 1998 to December 2003, he studied Bachelor of Science (B.Sc.) programme in Statistics at the Department of Statistics, School of Physical Sciences, Shahjalal University of Science & Technology, Sylhet, Bangladesh. He stood first in order of merit among all graduates of the Department of Statistics. From January 2004 to August 2005, he studied Master of Science (M.Sc.) programme in Statistics at the same university and from October 2008 to September 2010, he also studied Biostatistics programme with VLIR-UOS scholarship at the Center for Statistics, Hasselt University, Belgium. As of January 2011 to December 2014, he worked as a PhD candidate at the Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, The Netherlands, under supervision of Prof. Dr. A. de Boer, Prof. Dr. Kit C.B. Roes, Dr. R.H.H. Groenwold and Dr. O.H. Klungel.

Jamal started his academic career in February 2006, joining as a Lecturer in the Department of Statistics at Shahjalal University of Science & Technology, Sylhet, Bangladesh, and in November 2008, he was promoted to the Assistant Professor. During his PhD study, he was in study leave from his job in Bangladesh. As of January 2015, he is going to join his previous position (Assistant Professor) in Bangladesh.