Medical Decision Making
2020, Vol. 40(5) 582–595

DOI: 10.1177/0272989X20937257
journals.sagepub.com/home/mdm

A Systematic Review of Methods Used for Confounding Adjustment in Observational Economic Evaluations in Cardiology Conducted between 2013 and 2017

Jason R. Guertin, Blanchard Conombo, Raphaël Langevin, Frédéric Bergeron, Anne Holbrook, Brittany Humphries, Alexis Matteau, Brian J. Potter, Christel Renoux, Jean-Éric Tarride, and Madeleine Durand

Background. Observational economic evaluations (i.e., economic evaluations in which treatment allocation is not randomized) are prone to confounding bias. Prior reviews published in 2013 have shown that adjusting for confounding is poorly done, if done at all. Although these reviews raised awareness on the issues, it is unclear if their results improved the methodological quality of future work. We therefore aimed to investigate whether and how confounding was accounted for in recently published observational economic evaluations in the field of cardiology. **Methods.** We performed a systematic review of PubMed, Embase, Cochrane Library, Web of Science, and PsycInfo databases using a set of Medical Subject Headings and keywords covering topics in "observational economic evaluations in health within humans" and "cardiovascular diseases." Any study published in either English or French between January 1, 2013, and December 31, 2017, addressing our search criteria was eligible for inclusion in our review. Our protocol was registered with PROSPERO (CRD42018112391). **Results.** Forty-two (0.6%) out of 7523 unique citations met our inclusion criteria. Fewer than half of the selected studies adjusted for confounding (n = 19 [45.2%]). Of those that adjusted for confounding, propensity score matching (n = 8 [42.1%]) and other matching-based approaches were favored (n = 8 [42.1%]). Our results also highlighted that most authors who adjusted for confounding rarely justified their methodological choices. **Conclusion.** Our results indicate that adjustment for confounding is often ignored when conducting an observational economic evaluation. Continued knowledge translation efforts aimed at improving researchers' knowledge regarding confounding bias and methods aimed at addressing this issue are required and should be supported by journal editors.

Keywords
cardiology, confounding adjustment, economic evaluations, observational studies, systematic review

Date received: December 18, 2019; accepted: June 2, 2020

Economic evaluations in health aim to examine the incremental cost and incremental effectiveness of a health technology over 1 or many other health technologies. The vast majority of these are based on randomized controlled trials that evaluated the clinical effectiveness of health technologies.[1] However, observational studies, which do not randomize treatment allocations, are sometimes used to evaluate the effectiveness of health technologies. Such studies may also serve as the basis for an economic evaluation of the health technologies (hereby referred to as observational economic evaluations). Although these studies may be conducted prior to the reimbursement of a given

Corresponding Author
Jason R. Guertin, Centre de recherche du CHU de Québec–Université Laval, Axe Santé des populations et pratiques optimales en santé, Hôpital du Saint-Sacrement, 1050 chemin Sainte-Foy, local J1-09B, Québec, G1S 4L8, Canada (jason.guertin@fmed.ulaval.ca).

technology, they are particularly useful when reassessing previously reimbursed technologies as the results they highlight the true economic value of these technologies in the real-world setting.

Among other caveats (such as selection and information bias), observational studies are prone to confounding bias.[2–4] Adequately adjusting for this bias is fundamental when attempting to assess the true economic values of the health technologies being evaluated.[5,6] Methods traditionally used in epidemiology and economics to adjust for confounding bias (e.g., regression-based approaches, instrumental variables, propensity score [PS] methods) may also be used within observational economic evaluations.[5,7,8]

Despite the number of methods that can be used to adjust for confounding within observational economic evaluations, concerns have been raised about the quality of the confounding adjustment in such studies. For example, in a review of 43 observational economic evaluations published between 1990 and 2010 aimed at examining how confounding bias was adjusted for within these studies, Rovithis[4] noted that only 9% of selected studies provided detailed justification for the methods they used. In parallel to Rovithis, Kreif et al.[3] also examined this question in a distinct systematic review of 81 economic evaluations based on observational data (i.e., selected studies included any economic evaluation that used observational data to assess the cost component [1%], the effectiveness component [32%], or both components [67%]) published between 2000 and 2011.[3] In their review, Kreif et al. showed that at least 80% of selected studies did not assess the basic assumptions of the methods they used. Although a worrisome conclusion, seeing as both systematic reviews included only studies that adjusted for confounding, their results likely underestimated the scope of this issue. Faced with such results and to improve future work, Kreif et al. proposed a critical appraisal tool of statistical methods used in observational economic evaluations, which has since gained strong endorsement in the field.[9]

Although we believe that these 2 systematic reviews will have improved the methodological quality of published observational economic evaluations, no study has examined their impact on the field. Therefore, the primary objective of this study was to systematically review the methodological quality of recently published observational economic evaluations. Specifically, we asked 1) What proportion of studies used any method to adjust for confounding? and, when studies adjusted for confounding, 2) What was the quality of the confounding adjustment?

Methods

This systematic review has been registered at PROSPERO (CRD42018112391). Reporting of this systematic review followed the Preferred Report Item for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.[10,11]

Inclusion and Exclusion Criteria for Selecting Citations

A review of the literature published between January 1, 2013, and December 31, 2017, was conducted. We limited our systematic review to published cost-effectiveness, cost-utility, or cost-benefit analyses in the field of cardiology through the use of the "cardiovascular diseases" Medical Subject Headings (we refer readers to the search strategy presented in the Supplementary Table S1 for the breadth and depth of the coverage of the Medical Subject

Department of Social and Preventive Medicine, Université Laval, Quebec City, Canada (JRG, BC); Axe Santé des populations et pratiques optimales en santé, Centre de recherche du CHU de Québec–Université Laval, Quebec City, Canada (JRG, BC); Department of Economics, McGill University, Montreal, Canada (RL); Université Laval, Quebec City, Canada (FB); Division of Clinical Pharmacology and Toxicology, Department of Medicine, McMaster University, Hamilton, Canada (AH); Department of Health Evidence and Impact, McMaster University, Hamilton, Canada (AH, BH, JET); Department of Medicine, Université de Montréal, Montreal, Canada (AM, BJP, MD); Centre de recherche du Centre Hospitalier de l'Université de Montréal, Montreal, Canada (AM, BJP, MD); Centre Hospitalier de l'Université de Montréal, Montreal, Canada (AM, BJP, MD); McGill University, Montreal, Canada (CR); Centre for Health Economics and Policy Analysis, McMaster University, Hamilton, Canada (JET); Department of Economics; McMaster University, Hamilton, Canada (JET); Programs for Assessment of Technology in Health (PATH), The Research Institute of St. Joe's Hamilton, St. Joseph's Healthcare Hamilton (CR, JET); McMaster Chair in Health Technology Management, McMaster University, Hamilton, Canada (CR, JET).
The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article. The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was presented in abstract form at the Canadian Association for Population Therapeutics 2019 Annual Conference (Toronto, Canada) and at the 35th International Conference on Pharmacoepidemiology and Therapeutic Risk Management (Philadelphia, United States of America). Financial support for this study was provided entirely by a grant from the Canadian Institutes of Health Research (grant PJT-156292). This study also received an in-kind contribution from the Health and Social Services Systems, Knowledge Translation and Implementation Component of the Quebec SPOR-SUPPORT Unit. JR Guertin, BJ Potter, C Renoux, and M Durand are all research scholars of the Fonds de recherche du Québec en Santé. The funding agreements ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report.

Headings). We adapted the SIGN search filter on observational studies for our 5 databases.[12]

At least 2 reviewers assessed the following inclusion and exclusion criteria when screening eligible studies. To be included, both the cost and effectiveness components of all treatment groups were required to be fully based on observational data. We also included only peer-reviewed original articles examining human subjects and published in either the English or French language.

Exclusion criteria included solely using observational data to extrapolate results beyond the observation period (e.g., a model-based economic evaluation extrapolating short-term observed data to a lifetime model horizon without providing the short-term results would be excluded from this review). Other exclusion criteria included studies that did not compare at least 2 different interventions and studies that only stratified economic results based on an endogenous factor (i.e., a single-arm study in which the authors stratified the population based on a baseline sociodemographic or clinical characteristic and examined the incremental cost and effectiveness between the 2 patient strata such as in diabetics versus non diabetics patients).

Search Strategy

Our search strategy was designed in collaboration with an experienced research librarian (F.B.). The final search strategies were designed for PubMed, Embase, Cochrane Library, Web of Science, and PsycInfo databases (the search strategies used in the 5 databases are provided in the Supplementary Table S1).

Study Selection

Results from the search strategies were uploaded into a single EndNote X8 library. We used the "Find Duplicate" command within the EndNote program to eliminate duplicates citations identified within the 5 databases. Two independent reviewers (B.C. and R.L.) screened all remaining citations' titles and abstracts to determine eligibility; in case of disagreement at this stage, a third reviewer (J.R.G.) was consulted. Full text of citations deemed to be eligible for inclusion within the systematic review were independently reviewed by pairs of reviewers (J.R.G., B.C., R.L.) for final inclusion, with each reviewer reviewing two-thirds of eligible citations; disagreements at this stage were settled by the third reviewer. Selected full texts were included in the systematic review.

Data Collection

Data abstraction using a standardized form was performed independently by 3 independent reviewers (J.R.G., B.C., R.L.) working in pairs. In case of discrepancy, consensus was reached through discussion or through the involvement of the third reviewer.

Data Items

We extracted data pertaining to study characteristics (year of publication, journal type, jurisdiction of the analysis, study design, intervention and comparator types, sample sizes), design of the economic evaluation (format of the economic results [e.g., benefit-cost ratio, incremental cost-effectiveness ratio], inclusion of a cost-effectiveness acceptability curve and/or frontier), analysis plan (confounding adjustment method used, method used to account for missing data, extrapolation of results beyond the observed data), and the general conclusion of the main text regarding the cost-effectiveness of the study intervention(s) versus the comparator(s).

Analyses

We reported on the presence of confounding adjustment methods within all studies and, when present, assessed their quality using the Kreif et al. critical appraisal tool.[3] Briefly, the critical appraisal tool contains 6 questions and examines 3 broad categories of statistical methods that can be used to adjust for confounding bias: 1) instrumental variables, 2) regression-based methods, and 3) matching-based methods. Of note, these categories were defined by the authors in their publication, and some groupings may lead to confusion for certain readers, for example, inclusion of the "difference-in-difference method" in the "regression-based methods" or the "inverse probability of treatment weighting (IPTW) based on the PS" and "regression on the PS" (i.e., using the estimated PS as a covariate in a regression model) in the "matching-based methods." In our review, we chose to differentiate the difference-in-difference approach from other regression methods and differentiated the IPTW based on the PS from other PS matching methods but still evaluated the quality of confounding adjustment using the recommended questions. In addition, and unlike the recommendation in the critical appraisal tool, in studies in which multiple confounding adjustment methods were used, all were assessed independently.

We also examined the quality of the reporting in the selected studies of the items suggested by the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement.[13] Although the CHEERS statement identifies 24 items, those that were deemed irrelevant to single study-based observational economic evaluations (i.e., items 11b, 13b, 15, 16, 18, and 20b) were ignored; the presence or absence of the remaining 21 items was individually assessed within each selected study. All results were presented using descriptive statistics.

The various public funding agencies supporting this project or the authors had no role in the study.

Results

Our search strategy identified a total of 7523 unique citations in cardiology that were manually screened by at least 2 reviewers. After screening the titles and abstracts, 6970 citations (92.6%) were excluded, and the remaining full texts were screened by reviewers for final inclusion. Following this step, 42 unique citations were deemed eligible for inclusion in our review.[14–55] As shown in Figure 1 detailing the screening process, most of the full texts screened (n = 324 [63.4%]) were excluded because they were not a full economic evaluation.

Study Characteristics and Methodological Designs

Summary statistics of the identified observational economic evaluations are shown in Table 1 (study-specific characteristics are provided in the Supplementary Table S2). Briefly, of the 42 identified studies in the field of cardiology, 16.7% were published in 2013, 40% were published in 2014–15 and 43% in 2016–17. Most studies were conducted in a European (40%) or North American (21%) jurisdiction. Approximately two-thirds of the studies were based on claims (31%) or charts review (33%), while the remaining were based on nonrandomized trial data (19%) or registries (17%). Consequently, most studies used a historical cohort (also commonly referred to as retrospective) study design (e.g., 60%). Of note, although most studies compared 2 interventions of the same type (e.g., a pharmaceutical agent being compared with a different pharmaceutical agent), six (14%) of the 42 selected studies compared 2 different types of interventions (e.g., a patient management strategy to a medical device). Most results were presented as incremental cost-effectiveness ratios (86%).

Choice and Quality of the Confounding Adjustment Method Used

Table 2 summarizes the results of our assessment based on the Kreif et al.[3] critical appraisal tool of the choice and quality of the confounding adjustment methods used by the authors of the 42 selected studies (study-specific results are provided in Supplementary Table S3). Fewer than half of the studies (45%) reported adjusting for confounding. Among those that did report adjusting for confounding, most studies (84%) reported using a matching-based method. Of the 16 studies that used a matching-based method to adjust for confounding, most used either PS matching (50%) or covariate-based matching (44%), while a single study (6%) did not define what matching method was used. In addition, our results also highlight that studies that adjusted for confounding with the use of a matching-based method used it as the sole confounding adjustment method (69%) or in combination with a second method (31%). Remaining studies used either regression-based methods (26%), differences-in-differences (5%), IPTW based on the PS (5%), or a self-controlled study design (5%), and no study used an instrumental variable.

Few studies justified their methodological choices. Only 1 study (5%) assessed the “no unobserved confounding” assumption (question 1a of the critical appraisal tool). Such an observation does not stipulate that unobserved confounding is present but rather that authors generally did not report on the potential presence of unobserved confounding nor its potential impact on the results.

Similarly, no study assessed whether the distributions of the baseline covariates overlapped between the treatment groups (question 2 of the critical appraisal tool). Given that lack of overlap in the treatment groups’ baseline covariates can lead to biased results, we were unable to state if selected studies complied with this assumption or if additional steps should have been taken by the authors to ensure that baseline covariates within the analyzed study groups did overlap (e.g., restrict the analysis to a subset of the study population).[56]

Of the 5 studies (26%) that used regression-based methods and the 1 study (5%) that used the difference-in-difference approach, most did not assess the specification of the regression model for the health outcome (83%) nor for the costs (67%). For example, studies did not assess the relative fit of alternative models when reporting their results.

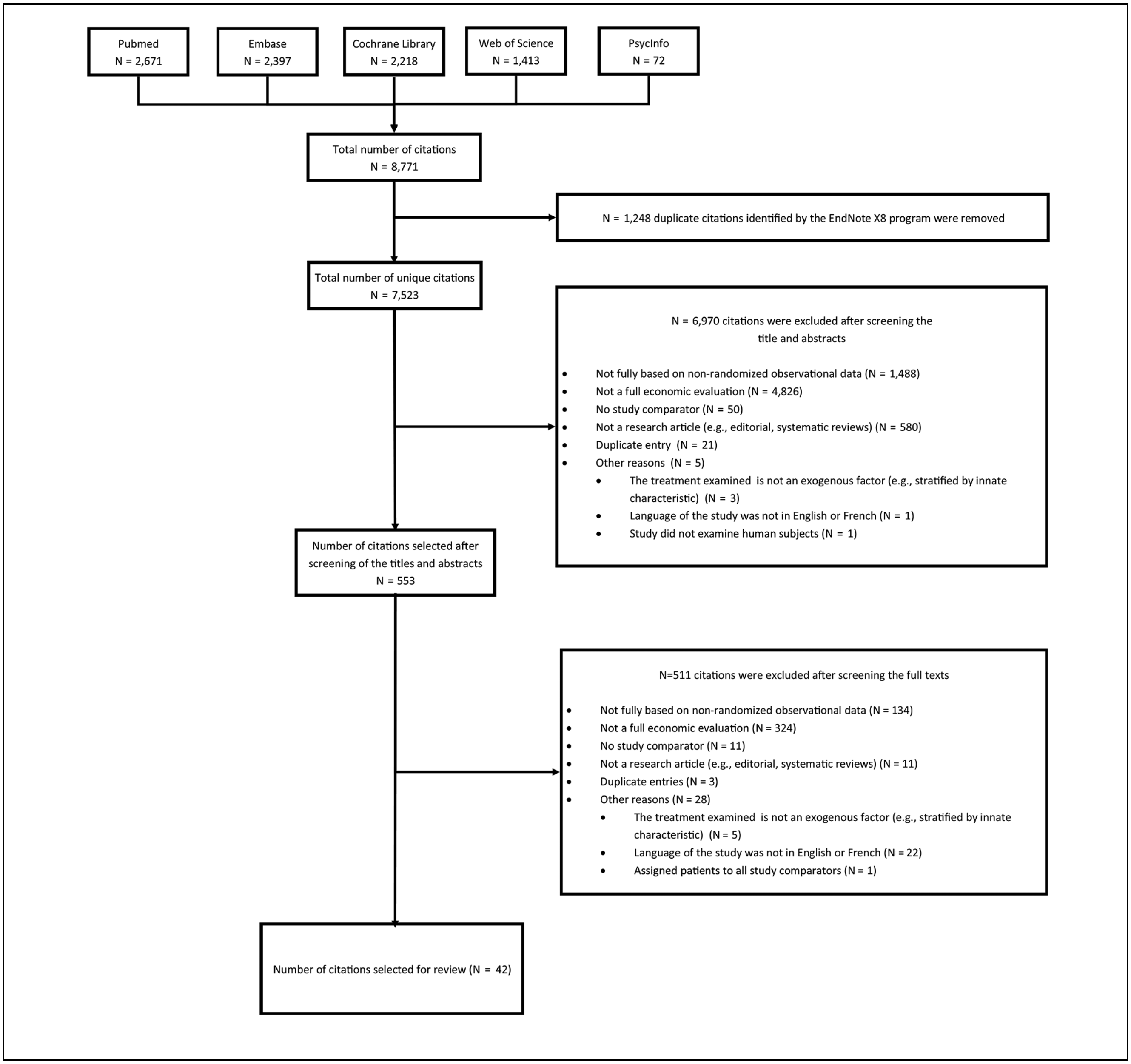

Figure 1 Flowchart of the citation selection process.

For the 16 studies that used matching-based methods, most studies (63%) only partially assessed covariate balance after applying the matching method (e.g., with the use of standardized differences) while the others (38%) did not. We were therefore unable to assess if residual confounding remains despite the matching procedures that were conducted.

Quality of the Item Reporting as Assessed with the CHEERS Statement

The proportion of studies reporting each of the items identified within the CHEERS statement is provided in Table 3 (Supplementary Table S4 provides the study-specific results). Although none of the studies acknowledged the CHEERS statement, selected studies reported

Table 1 Basic Characteristics of the Selected Studies

Year of publication	
2013	7 (16.7)
2014	11 (26.2)
2015	6 (14.3)
2016	4 (9.5)
2017	14 (33.3)
Journal type	
Health economics	4 (9.5)
Statistics	0 (0.0)
Medical	36 (85.7)
Other	2 (4.8)
Jurisdiction of the study[a]	
China	4 (9.5)
India	3 (7.1)
Spain	4 (9.5)
United Kingdom	5 (11.9)
United States of America	9 (21.4)
Undefined	2 (4.8)
Other	15 (35.7)
Data source	
Claims based	13 (30.9)
Nonrandomized trial based	8 (19.0)
Chart review based	14 (33.3)
Registry based	7 (16.7)
Study design	
Concurrent cohort	15 (35.7)
Pre-post design	1 (2.4)
Historical cohort	25 (59.5)
Other	1 (2.4)
Authors reported on the presence or absence of missing data	
Yes	8 (19.0)
No	34 (80.9)
Method used to account for missing data (n = 8 who reported on the presence or absence of missing data)	
No missing data	1 (12.5)
Complete case analysis	3 (37.5)
Simple imputation	1 (12.5)
Multiple imputation	1 (12.5)
Did not report how they accounted for the missing data in their analysis	2 (25.0)
Interventions being compared	
Standard of care	
Medical device	7 (16.7)
Pharmaceutical agent	9 (21.4)
Prevention and screening	3 (7.1)
Patient management	18 (42.8)
Other	5 (11.9)
Comparator[b]	
Medical device	10 (23.8)
Pharmaceutical agent	11 (26.2)
Prevention and screening	3 (7.1)
Patient management	13 (30.9)
Other	5 (11.9)
Sample size of the standard of care[c]	
25 or less	5 (11.9)
26 to 50	7 (16.6)
51 to 100	12 (28.6)
101 to 500	9 (21.4)
501 or more	7 (16.6)
Not reported	2 (4.8)
Sample size of the comparator[c,d]	
25 or less	4 (9.5)
26 to 50	8 (19)
51 to 100	10 (23.8)
101 to 500	10 (23.8)
501 or more	11 (26.2)
Not reported	1 (2.4)
Format of the economic result[e]	
ICER	36 (85.7)
Benefit-cost ratio	5 (11.9)
NMB	1 (2.4)
Provided results using a cost-effectiveness acceptability curve	
Yes	6 (14.3)
No	36 (85.7)
Provided results using a cost-effectiveness acceptability frontier	
Yes	0 (0.0)
No	42 (100.0)
Authors extrapolated the cost, the effectiveness, or both beyond the observed period	
Yes	3 (7.1)
No	39 (92.6)
General conclusion of the main text regarding the economic evaluation comparing the comparator to the standard of care	
Cost-effective	23 (54.8)
Borderline cost-effective	0 (0.0)
Not cost-effective	1 (2.4)
Dominant/dominated	7 (16.7)
Undefined	11 (26.2)

Results are presented as n (%). ICER = incremental cost-effectiveness ratio; NMB = net monetary benefit.
[a]Only countries listed as the jurisdiction of at least 3 studies are listed in this table.
[b]Some studies included >1 comparator, but all of these were of the same type and were therefore collapsed together in this table.
[c]n = 7 studies stratified their population into multiple subpopulations; categorization for these studies was based on the smallest strata.
[d]n = 5 studies included multiple comparators; categorization for these studies was based on the smallest sample size for the comparator group.
[e]Sum exceeds 100% as 1 study provided results as both incremental cost-effectiveness ratio and net monetary benefits.

an average of 14 of the 21 items identified in the checklist (range: 6–19).

Discussion

To the best of our knowledge, and unlike previous reviews that selected studies based on known confounding adjustment methods,[3,4] this is the first systematic review to examine how confounding adjustment was

Table 2 Quality of the Adjustment for Confounding Bias as Based on the Kreif et al.[3] Critical Appraisal Tool

Q0a. Did the authors consider adjusting for confounding?	
Yes	19 (45.2)
No	23 (54.8)
Only among those that adjusted for confounding ($n = 19$)	
What confounding adjustment method was used?[a,b]	
Difference-in-difference	1 (5.3)
Instrumental variable	0 (0.0)
Inverse probability of treatment weighting based on the propensity score	1 (5.3)
Matching	16 (84.2)
Regression	5 (26.3)
Self-controlled design	1 (5.3)
What type of matching method was used, if any ($n = 16$)?	
Covariate based	7 (43.8)
Propensity score	8 (50.0)
Undefined	1 (6.3)
Q1a. Did the study assess the "no unobserved confounding" assumption?[c]	
Yes	1 (5.3)
Partially	0 (0.0)
No	18 (94.7)
Q2. Did the study assess whether the distributions of the baseline covariates overlapped between the treatment groups?[c]	
Yes	0 (0.0)
Partially	0 (0.0)
No	19 (100.0)
Q3. Did the study assess the specification of the regression model for ($n = 6$)?[c,d]	
1) Health outcome	
Yes	1 (16.7)
Partially	0 (0.0)
No	5 (83.3)
2) Costs	
Yes	0 (0.0)
Partially	2 (33.3)
No	4 (66.7)
Q4. Was covariate balance assessed after applying a matching method ($n = 16$)?[c,e]	
Yes	0 (0.0)
Partially	10 (62.5)
No	6 (37.5)
Q5: Did the study consider structural uncertainty arising from the choice or specification of the statistical method for addressing selection bias?[c, f]	
Yes	1 (5.3)
Partially	1 (5.3)
No	17 (89.5)

Results are presented as n (%).
[a]The critical appraisal tool has a distinct question regarding the validity of the instrumental variable, when used. However, because none of the selected study adjusted for confounding using an instrumental variable, this question was dropped from the table.
[b]$n = 5$ studies used 2 confounding adjustment methods.
[c]A detailed description of each question in their critical appraisal tool and of the associated answer keys are provided elsewhere.[3]
[d]In accordance with the recommendations by Kreif et al.,[3] the quality of the difference-in-difference adjustment was assessed within question 3.
[e]One study matched patients on their propensity score using matching weights derived from the inverse probability of treatment. We combined the assessment of the methods for this study.
[f]In cases in which studies used more than 1 confounding adjustment method, Kreif et al.[3] would consider these alternative methods as ways to assess the structural uncertainty of the chosen statistical method. Since in the 5 studies that used 2 methods we assessed each individually, we did not include these alternative methods as ways to assess the structural uncertainty of the primary methods.

conducted within observational economic evaluations without this selection criterion. By doing so, we identified 42 observational economic evaluations published between 2013 and 2017 in the field of cardiology.[14–55] Our review of these studies identified that more than half did not attempt to adjust for confounding. Such results are surprising since the importance of correctly adjusting for confounding has been the focus of much work in

Table 3 Quality of the Reporting of Items Identified within the CHEERS Statement

Section/Item[a,b]	Item Number	Recommendation	Number of Publications Reporting the Item (%)
Title and abstract			
Title	1	Identify the study as an economic evaluation, or use more specific terms such as "cost-effectiveness analysis" and describe the interventions compared	12 (28.6)
Abstract	2	Provide a structured summary of objectives, perspective, setting, methods (including study design and inputs), results (including base-case and uncertainty analyses), and conclusions	27 (64.3)
Introduction			
Background and objectives	3	Provide an explicit statement of the broader context for the study; present the study question and its relevance for health policy or practice decisions	42 (100.0)
Methods			
Target population and subgroups	4	Describe characteristics of the base-case population and subgroups analyzed including why they were chosen	41 (97.6)
Setting and location	5	State relevant aspects of the system(s) in which the decision(s) need(s) to be made	39 (92.9)
Study perspective	6	Describe the perspective of the study and relate this to the costs being evaluated	20 (47.6)
Comparators	7	Describe the interventions or strategies being compared and state why they were chosen	37 (88.1)
Time horizon	8	State the time horizon(s) over which costs and consequences are being evaluated and say why appropriate	21 (50.0)
Discount rate	9	Report the choice of discount rate(s) used for costs and outcomes and say why appropriate	8 (19.0)
Choice of health outcomes	10	Describe what outcomes were used as the measure(s) of benefit in the evaluation and their relevance for the type of analysis performed	36 (85.7)
Measurement of effectiveness	11a	Single study-based estimates: describe fully the methods used for identification of included studies and synthesis of clinical effectiveness data	35 (83.3)
Measurement of preference-based outcomes	12	If applicable, describe the population and methods used to elicit preferences for outcomes	14 (70.0)[c]
Estimating resources and costs	13a	Single study-based economic evaluation: Describe approaches used to estimate resource use associated with the alternative interventions. Describe primary or secondary research methods for valuing each resource item in terms of its unit cost. Describe any adjustment made to approximate to opportunity costs.	35 (83.3)
Currency, price date, and conversion	14	Report the dates of the estimated resource quantities and unit costs; describe methods for adjusting estimated unit costs to the year of reported costs if necessary; describe methods for converting costs into a common currency base and the exchange rate	16 (38.1)
Analytic methods	17	Describe all analytic methods supporting the evaluation—this could include methods for dealing with skewed, missing, or censored data; extrapolation methods; methods for pooling data; approaches to validate or make adjustments (e.g., half-cycle corrections to a model); and methods for handling population heterogeneity and uncertainty	35 (83.3)
Results			
Incremental costs and outcomes	19	For each intervention, report mean values for the main categories of estimated costs and outcomes of interest, as well as mean differences between the comparator groups; if applicable, report incremental cost-effectiveness ratios	39 (92.9)

(continued)

Table 3 (continued)

Section/Item[a,b]	Item Number	Recommendation	Number of Publications Reporting the Item (%)
Characterizing uncertainty	20a	Single study-based economic evaluation: describe the effects of sampling uncertainty for estimated incremental cost, incremental effectiveness, and incremental cost-effectiveness, together with the impact of methodological assumptions (such as discount rate, study perspective)	19 (45.2)
Characterizing heterogeneity	21	If applicable, report differences in costs, outcomes, or cost-effectiveness that can be explained by variations between subgroups of patients with different baseline characteristics or other observed variability in effects that are not reducible by more information	12 (28.6)
Discussion			
Study findings, limitations, generalizability, and current knowledge	22	Summarize key study findings and describe how they support the conclusions reached; discuss limitations and the generalizability of the findings and how the findings fit with current knowledge	36 (85.7)
Other			
Source of funding	23	Describe how the study was funded and the role of the funder in the identification, design, conduct, and reporting of the analysis; describe other nonmonetary sources of support	27 (64.3)
Conflicts of interest	24	Describe any potential for conflict of interest among study contributors in accordance with journal policy; in the absence of a journal policy, we recommend authors comply with International Committee of Medical Journal Editors' recommendations	38 (90.5)

[a]A detailed description of all of the items included in the CHEERS statement can be found elsewhere.[57]
[b]Items 15, 16, and 18 of the CHEERS checklist were ignored as they pertained to items considered irrelevant to the types of studies examined within the review.
[c]This item was not applicable in 22 of the 42 selected publications (52.3%). The relative value was calculated from the 20 remaining publications where this item was relevant.

epidemiology, economics, and health economics. However, regardless of the reason why confounding factors were not adjusted for in these studies, we can easily state that their unadjusted results are biased. Unfortunately, since confounding can bias both the economic component and the effectiveness component of the economic evaluation,[5,6] we were unable to evaluate if the conclusions of these studies would differ had they adjusted for confounding.

Quantitatively, even though we restricted our review to observational economic evaluations in the field of cardiology, many of our results regarding the studies that did adjust for confounding are aligned with those found in past systematic reviews on the subject that did not impose such a restriction. Like others,[3,4] we show that authors who adjusted for confounding favored regression and matching-based methods (Table 2). However, unlike Kreif et al.,[3] who found that authors favored regression-based methods (51%), including the difference-in-difference method, over covariate matching (25%) and all PS methods (22%), we showed that PS methods (42%) and covariate matching (37%) were favored over regression-based methods (26%) and the difference-in-difference method (5%). Our results are therefore more aligned with those obtained by Rovithis[4] that showed that PS methods (49%) and regression-based methods (28%) were favored over others. Of note, although no document has identified whether PS methods are superior to other commonly available methods when adjusting for confounding in observational economic evaluation, this result probably reflects the increased uptake of PS methodology in other fields[58,59] and the relative ease with which PS methods may be used in these types of studies.[8]

Despite these similarities with regard to the choice of confounding methods being used by authors, our results suggest that the quality of the confounding adjustment in recently published studies in cardiology tends to be worse than in those selected by Kreif et al.[3] within their review (Table 2). For example, none of our selected studies assessed the "no unobserved confounding" assumption (i.e., Q1a in the Kreif et al. critical appraisal tool), whereas 23% studies at least partially assessed this assumption in their review. Although it is unclear why the quality of confounding adjustment would be poorer in the studies we selected, one explanation could reside in the fact that Kreif et al. selected studies using a broader selection criterion (i.e., identified economic evaluations were eligible as long as at least 1 component of the economic evaluation was based on observational data). Unfortunately, results limited to the subset of studies in which both the cost and the effectiveness components were based on observational data are not available.

In a broader context, our review raises 2 important issues regarding the value of and the choice of methods used to adjust for confounding in observational economic evaluations. First, as we previously noted, we were unable to predict the impact of confounders in the studies that only provided unadjusted results nor if adjusted results would lead to different conclusions regarding the value of the study interventions. One particular solution that is commonly used in epidemiology is to examine the theoretical impact that an unadjusted confounder could have on the observed association.[60] To the best of our knowledge, this approach has never been extended to the context of observational economic evaluations. Continued work focusing on this area of research is warranted. Second, despite the value of having a short and quick critical appraisal tool containing only 6 questions to asses the quality of confounding adjustment, our review hints to the limits of assessing the validity of the confounding adjustment using such a limited number of questions, especially considering the ongoing developments in this area. We recognize that all of the methods used within the selected studies were originally considered by Kreif et al.[3] However, many new confounding adjustment methods have been developed in recent years, for example, semiautomated techniques (e.g., high-dimensional PS methods[61]), G methods,[62,63] or double robust methods that can naturally be combined with machine-learning methods (e.g., the use of Super Learner within targeted maximum likelihood estimation[64,65]). We refer interested readers to various methodological papers and primers describing the strengths and limits of some of these methods in the context of adjustment for confounding.[59,64,66–71] Future work aimed at revising previous tools to better integrate these novel methods and comparing these various methods to each other in the context of observational economic evaluations is also needed.

In light of these issues, we make the following 2 recommendations to be considered in future good practice guidelines. First, we suggest that authors report both the unadjusted and adjusted results of their observational economic evaluations, as the difference between the 2 results may hint to the potential impact of confounding in the examined setting. Second, we advise researchers to further disclose and justify the confounding adjustment method they use when conducting an observational economic evaluation. We believe that such a recommendation could be facilitated by the inclusion of additional confounding adjustment-related items in a revised version of the CHEERS statement or in a novel reporting guideline specific to observational economic evaluations; these items could be inspired by those present in the STrenghtening the Reporting of Observational studies in Epidemiology (STROBE) statement.[72,73] We recognize that our results show that many items in the CHEERS statement were not reported within our selected studies (Table 3) and that only 2 were published in journals that previously endorsed this reporting guideline (i.e., *Value in Health* and the *Journal of Medical Economics*). That being said, previous work by others have shown that reporting of at least certain items included in reporting guidelines improved following their publication.[74–76] Although we believe that adding a critical appraisal tool, like the one proposed by Kreif et al.,[3] could improve the reporting as well as the quality of the confounding adjustment, revisions to these tools are required (e.g., adding items specific to novel methods) before formally recommending their adoption by journals.

Our study has several strengths and limitations. The primary strength of our review of observational economic evaluations in cardiology resides in the fact that, unlike others,[3,4] we did not restrict our review to only published economic evaluations that stated having adjusted for confounding bias. By doing so, we were able to identify that more than half (54.8%) of identified observational economic evaluations in cardiology did not adjust for confounding bias. That being said, to do so, we created a search strategy that favored sensitivity over specificity. To compensate for the high sensitivity and low specificity of this search strategy, we chose to limit our review to manuscripts published between 2013 and 2017 and did not examine other clinical areas to maintain its feasibility. Although we cannot exclude the risk that our results do not reflect the broader current state of the field, the clinical area we chose to examine remains one of the most important in terms of volume in PubMed. As such, we believe our results likely also reflect the current state of the literature in terms of confounding adjustment within observational economic evaluations in other clinical areas. Another noticeable strength of our work compared with previous reviews[3,4] is the fact that the article screening and selection as well as the data extraction were always conducted by at least 2 independent members of our team. In terms of limitations, our assessment of the quality of the confounding adjustment was limited by the quality of the reporting within the study. As such, we cannot exclude the fact that authors who did control for confounding did account for the underlying assumptions associated with these methods but did not report them. However, this limit should not undermine our result regarding the high proportion of studies that were limited to unadjusted results. An additional limitation of our work concerns the fact that we only examined a fraction of the types of economic evaluations based on observational data, that is, only those in which both the cost and effectiveness components were based solely on observational data. There are, however, numerous other types of economic evaluations based on observational data, including those that use observational data only for 1 of the 2 components of the economic evaluation and model-based economic evaluations extrapolating observational data. Although we are unable to state whether our results regarding the quality of the confounding adjustment extends to these alternative types of studies, our conclusions regarding the importance of adjusting for confounding remain valid. Future work is also needed to better examine the limits of using observational data in these other types of studies.

In conclusion, our results highlight the fact that more than half of observational economic evaluations in cardiology, published between 2013 and 2017, did not adjust for potential confounders and that, in those that did, the methodological choices were rarely justified by the authors. Despite previous knowledge translation efforts, additional knowledge translation efforts are warranted. One potential solution could reside in including confounding adjustment-related items in the upcoming revision of the CHEERS statement[77] and/or producing an extension of the CHEERS statement specific to observational economic evaluations that would state the importance of reporting on the method chosen to address confounding bias in studies and on the risk for residual confounding.

Acknowledgments

We thank Nathalie Rheault, information specialist, and Hervé Tchala Vignon Zomahoun, scientific coordinator, from the "Health and Social Services Systems, Knowledge Translation and Implementation" Component of the Quebec SPOR-SUPPORT Unit, who supported us in developing the literature search strategy and identifying the unique citations for the study selection process.

ORCID iDs

Jason R. Guertin https://orcid.org/0000-0003-1718-5307
Blanchard Conombo https://orcid.org/0000-0001-6636-8164

Supplemental Material

Supplementary material for this article is available on the *Medical Decision Making* Web site at http://journals.sagepub.com/home/mdm.

References

1. Ramsey SD, Willke RJ, Glick H, et al. Cost-effectiveness analysis alongside clinical trials II: an ISPOR Good Research Practices Task Force report. *Value Health*. 2015;18(2):161–72.
2. Deidda M, Geue C, Kreif N, Dundas R, McIntosh E. A framework for conducting economic evaluations alongside natural experiments. *Soc Sci Med*. 2019;220:353–61.
3. Kreif N, Grieve R, Sadique MZ. Statistical methods for cost-effectiveness analyses that use observational data: a critical appraisal tool and review of current practice. *Health Econ*. 2013;22(4):486–500.
4. Rovithis D. Do health economic evaluation using observational data provide reliable assessment of treatment effects? *Health Econ Rev*. 2013;3:21.
5. Hoch JS, Briggs AH, Willan AR. Something old, something new, something borrowed, something blue: a framework for the marriage of health econometrics and cost-effectiveness analysis. *Health Econ*. 2002;11(5):415–30.
6. Guertin JR, Bowen JM, De Rose G, O'Reilly DJ, Tarride J-E. Illustration of the impact of unmeasured confounding within an economic evaluation based on nonrandomized data. *MDM Policy Pract*. 2017;2(1):1–11.
7. Indurkhya A, Mitra N, Schrag D. Using propensity scores to estimate the cost-effectiveness of medical therapies. *Stat Med*. 2006;25(9):1561–76.
8. Manca A, Austin PC. *Using Propensity Score Methods to Analyse Individual Patient Level Cost Effectiveness Data from Observational Studies*. York (UK): The University of York, Health Economics and Data Group Working Paper; 2008;8:20.
9. Drummond M, Sculpher M, Claxton K, Stoddart G, Torrance G. *Methods for the Economic Evaluation of Health Care Programmes*. 4th ed. Oxford (UK): Oxford University Press; 2015.
10. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Med*. 2009;6(7):e1000100.

11. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* 2009;6(7): e1000097.
12. Scottish Intercollegiate Guidelines. Search filters: observational studies. Edinburgh (UK): SIGN; 2019. Available from: https://www.sign.ac.uk/search-filters.
13. Husereau D, Drummond M, Petrou S, et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. *BMJ*. 2013;346:f1049.
14. Abdel-Rahman F, Tuffaha HW, Sharma S, et al. GCSF with or without chemotherapy compared to Plerixafor with GCSF as salvage mobilization regimen in patients with multiple myeloma and lymphoma: collection effectiveness and cost effectiveness analysis. *J Oncol Pharm Pract*. 2014;20(2):130–6.
15. Aitken EL, Jackson AJ, Hameed H, Chandramohan M, Kasthuri R, Kingsmore DB. Cephalic arch stenosis: angioplasty to preserve a brachiocephalic fistula or new brachiobasilic fistula? A cost-effectiveness study. *Ren Fail*. 2014; 36(10):1550–8.
16. Aljutaili M, Becker C, Witt S, et al. Should health insurers target prevention of cardiovascular disease? A cost-effectiveness analysis of an individualised programme in Germany based on routine data. *BMC Health Serv Res*. 2014;14:263.
17. Gopi B, Sushmitha C, Nikitha K, et al. Cost-effectiveness analysis in the management of stroke. *Asian J Pharm Clin Res*. 2017;10(7):127–30.
18. Behar JM, Chin HM, Fearn S, et al. Cost-effectiveness analysis of quadripolar versus bipolar left ventricular leads for cardiac resynchronization defibrillator therapy in a large, multicenter UK registry. *JACC Clin Electrophysiol*. 2017;3(2):107–16.
19. Chang AM, Ho JC, Yan BP, Yu CM, Lam YY, Lee VW. Cost-effectiveness of dabigatran compared with warfarin for stroke prevention in patients with atrial fibrillation—a real patient data analysis in a Hong Kong teaching hospital. *Clin Cardiol*. 2013;36(5):280–5.
20. Corso A, Mangiacavalli S, Cocito F, et al. Long term evaluation of the impact of autologous peripheral blood stem cell transplantation in multiple myeloma: a cost-effectiveness analysis. *PLoS One*. 2013;8(9):e75047.
21. de Lorenzo-Pinto A, Herranz-Alonso A, Cuellar-Basterrechea B, et al. Clinical and economic impact of a multidisciplinary intervention to reduce bleeding risk in patients with acute coronary syndrome. *Revista espanola de cardiologia*. 2017;70(10):825–31.
22. El-Saiedi SA, El Sisi AM, Mandour RS, Abdel-Aziz DM, Attia WA. Cost-effectiveness analysis of different devices used for the closure of small-to-medium-sized patent ductus arteriosus in pediatric patients. *Ann Pediatr Cardiol*. 2017;10(2):144–51.
23. Eseonu CI, Rincon-Torroella J, ReFaey K, Quinones-Hinojosa A. The cost of brain surgery: awake vs asleep craniotomy for perirolandic region tumors. *Neurosurgery*. 2017;81(2):307–14.
24. Flynt LK, Kenney RM, Zervos MJ, Davis SL. The safety and economic impact of cefazolin versus nafcillin for the treatment of methicillin-susceptible *Staphylococcus aureus* bloodstream infections. *Infect Dis Ther*. 2017;6(2):225–31.
25. Freeman PM, Protty MB, Aldalati O, et al. Severe symptomatic aortic stenosis: medical therapy and transcatheter aortic valve implantation (TAVI)-a real-world retrospective cohort analysis of outcomes and cost-effectiveness using national data. *Open Heart*. 2016;3(1):e000414.
26. Hlatky MA, Shilane D, Hachamovitch R, Dicarli MF, Investigators S. Economic outcomes in the Study of Myocardial Perfusion and Coronary Anatomy Imaging Roles in Coronary Artery Disease registry: the SPARC Study. *J Am Coll Cardiol*. 2014;63(10):1002–8.
27. Ho YL, Yu JY, Lin YH, et al. Assessment of the cost-effectiveness and clinical outcomes of a fourth-generation synchronous telehealth program for the management of chronic cardiovascular disease. *J Med Internet Res*. 2014; 16(6):e145.
28. Hsieh HM, Gu SM, Shin SJ, Kao HY, Lin YC, Chiu HC. Cost-effectiveness of a diabetes pay-for-performance program in diabetes patients with multiple chronic conditions. *PLoS One*. 2015;10(7):e0133163.
29. Jeong HS, Shin JW, Kwon HJ, et al. Cost benefits of rapid recanalization using intraarterial thrombectomy. *Brain Behav*. 2017;7(10):e00830.
30. Krenn L, Kopp C, Glogar D, et al. Cost-effectiveness of percutaneous coronary intervention with drug-eluting stents in patients with multivessel coronary artery disease compared to coronary artery bypass surgery five-years after intervention. *Catheter Cardiovasc Intervent*. 2014;84(7):1029–39.
31. Langabeer JR III, Champagne-Langabeer T, Alqusairi D, et al. Cost-benefit analysis of telehealth in pre-hospital care. *J Telemed Telecare*. 2017;23(8):747–51.
32. Lee S, Baek K, Chun K.Cost-effectiveness of drug-eluting vs. bare-metal stents in patients with coronary artery disease from the Korean National Health Insurance Database. *Yonsei Med J*. 2014;55(6):1533–41.
33. Lee VW, Cheng FW, Choi AY, Fong ST, Yu CM, Yan BP. Clinical, humanistic, and economic outcomes between drug-eluting stent (DES) and bare metal stent (BMS): 18-month follow-up study. *J Med Econ*. 2017;20(3):239–45.
34. Maeng DD, Pitcavage JM, Tomcavage J, Steinhubl SR. Can health insurance improve employee health outcome and reduce cost? An evaluation of Geisinger's employee health and wellness program. *J Occup Environ Med*. 2013; 55(11):1271–5.
35. Mafe JJ, Planelles B, Asensio S, et al. Cost and effectiveness of lung lobectomy by video-assisted thoracic surgery for lung cancer. *J Thorac Dis*. 2017;9(8):2534–43.
36. Malhotra AK, Goldberg SR, McLay L, et al. DVT surveillance program in the ICU: analysis of cost-effectiveness. *PLoS One*. 2014;9(9):e106793.

37. Nilsson S, Andersson A, Janzon M, Karlsson JE, Levin LA. Cost consequences of point-of-care troponin T testing in a Swedish primary health care setting. *Scand J Prim Health Care*. 2014;32(4):241–7.
38. Norton SP, Dickerson EM, Kulwin CG, Shah MV. Technology that achieves the Triple Aim: an economic analysis of the BrainPath approach in neurosurgery. *Clinicoecon Outcomes Res*. 2017;9:519–23.
39. Panca M, Cutting K, Guest JF. Clinical and cost-effectiveness of absorbent dressings in the treatment of highly exuding VLUs. *J Wound Care*. 2013;22(3):109–10, 12–8.
40. Parody E, Pedraza S, Garcia-Gil MM, Crespo C, Serena J, Davalos A. Cost-utility analysis of magnetic resonance imaging management of patients with acute ischemic stroke in a Spanish hospital. *Neurol Ther*. 2015;4(1):25–37.
41. Parvathy J, Philip TM, Abhijith K, Sreeni S, Panicker NK, Nambiar V. Fluoxetine vs venlafaxine: economic evaluation in post stroke depression. *Int J Pharm Sci Rev Res*. 2016;41(1):234–6.
42. Pattenden JF, Mason AR, Lewin RJ. Collaborative palliative care for advanced heart failure: outcomes and costs from the 'Better Together' pilot study. *BMJ Support Palliat Care*. 2013;3(1):69–76.
43. Peres J, Martins R, Alves JD, Valverde A. Rituximab in generalized myasthenia gravis: clinical, quality of life and cost–utility analysis. *Porto Biomed J*. 2017;2(3):81–5.
44. Petrov G, Kelle S, Fleck E, Wellnhofer E. Incremental cost-effectiveness of dobutamine stress cardiac magnetic resonance imaging in patients at intermediate risk for coronary artery disease. *Clin Res Cardiol*. 2015;104(5):401–9.
45. Pribadi FW, Dwiprahasto I, Thobari JA. Cost analysis of combination diuretic therapy with Ace-inhibitors to diuretic therapy without Ace-inhibitors in heart failure patients. *Bangladesh J Med Sci*. 2016;15(3):368–75.
46. Regueiro A, Bosch J, Martin-Yuste V, et al. Cost-effectiveness of a European ST-segment elevation myocardial infarction network: results from the Catalan Codi Infart network. *BMJ Open*. 2015;5(12):e009148.
47. Rychlik R, Kreimendahl F, Schnur N, Lambert-Baumann J, Dressler D. Quality of life and costs of spasticity treatment in German stroke patients. *Health Econ Rev*. 2016;6(1):27.
48. Spence MM, Makarem AF, Reyes SL, et al. Evaluation of an outpatient pharmacy clinical services program on adherence and clinical outcomes among patients with diabetes and/or coronary artery disease. *J Manag Care Spec Pharm*. 2014;20(10):1036–45.
49. Sultan S, Tawfick W, Hynes N. Cool excimer laser-assisted angioplasty (CELA) and tibial balloon angioplasty (TBA) in management of infragenicular arterial occlusion in critical lower limb ischemia (CLI). *Vasc Endovascular Surg*. 2013;47(3):179–91.
50. Tamilselvan T, Kumutha T, Lekshmi AV, James AC, Reji JS, Cheriyan N. Pharmacoeconomical evaluation of oral hypoglycemic agents for type-2 diabetes mellitus in a multispeciality hospital. *Int J Pharm Sci Res*. 2017;8(5).
51. Tsiachristas A, Burgers L, Rutten-van Molken MP. Cost-effectiveness of disease management programs for cardiovascular risk and COPD in The Netherlands. *Value Health*. 2015;18(8):977–86.
52. Wittayanukorn S, Westrick SC, Hansen RA, et al. Evaluation of medication therapy management services for patients with cardiovascular disease in a self-insured employer health plan. *J Manag Care Pharm*. 2013;19(5):385–95.
53. Yan X, Hu H-T, Liu S, Sun Y-H, Gao X. A pharmacoeconomic assessment of recombinant tissue plasminogen activator therapy for acute ischemic stroke in a tertiary hospital in China. *Neurol Res*. 2014;37(4):352–8.
54. Zhang H, Wang Y, Jiang ZM, et al. Impact of nutrition support on clinical outcome and cost-effectiveness analysis in patients at nutritional risk: a prospective cohort study with propensity score matching. *Nutrition*. 2017;37:53–9.
55. Zhang Z, Kolm P, Grau-Sepulveda MV, et al. Cost-effectiveness of revascularization strategies: the ASCERT study. *J Am Coll Cardiol*. 2015;65(1):1–11.
56. Petersen ML, Porter KE, Gruber S, Wang Y, van der Laan MJ. Diagnosing and responding to violations in the positivity assumption. *Stat Methods Med Res*. 2012;21(1):31–54.
57. Husereau D, Drummond M, Petrou S, et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS)—explanation and elaboration: a report of the ISPOR Health Economic Evaluation Publication Guidelines Good Reporting Practices Task Force. *Value Health*. 2013;16(2):231–50.
58. Borah BJ, Moriarty JP, Crown WH, Doshi JA. Applications of propensity score methods in observational comparative effectiveness and safety research: where have we come and where should we go? *J Comp Eff Res*. 2014;3(1):63–78.
59. Crown WH. Propensity-score matching in economic analyses: comparison with regression models, instrumental variables, residual inclusion, differences-in-differences, and decomposition methods. *Appl Health Econ Health Policy*. 2014;12(1):7–18.
60. Groenwold RH, Nelson DB, Nichol KL, Hoes AW, Hak E. Sensitivity analyses to estimate the potential impact of unmeasured confounding in causal research. *Int J Epidemiol*. 2010;39(1):107–17.
61. Schneeweiss S, Rassen JA, Glynn RJ, Avorn J, Mogun H, Brookhart MA. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. *Epidemiology*. 2009;20(4):512–22.
62. Naimi AI, Cole SR, Kennedy EH. An introduction to g methods. *Int J Epidemiol*. 2017;46(2):756–62.
63. Robins JM, Hernán MA. Estimation of the causal effects of time-varying exposures. *Longitudinal Data Analysis*. 2009;553:599.
64. Wyss R, Schneeweiss S, van der Laan M, Lendle SD, Ju C, Franklin JM. Using Super Learner prediction modeling to improve high-dimensional propensity score estimation. *Epidemiology*. 2018;29(1):96–106.

65. van der Laan MJ, Rubin D. Targeted maximum likelihood learning. *Int J Biostat*. 2006;2(1).
66. Austin PC. The relative ability of different propensity score methods to balance measured covariates between treated and untreated subjects in observational studies. *Med Decis Making*. 2009;29(6):661–77.
67. Inacio MC, Chen Y, Paxton EW, Namba RS, Kurtz SM, Cafri G. Statistics in brief: an introduction to the use of propensity scores. *Clin Orthop Relat Res*. 2015;473(8):2722–6.
68. Williamson E, Morley R, Lucas A, Carpenter J. Propensity scores: from naive enthusiasm to intuitive understanding. *Stat Methods Med Res*. 2012;21(3):273–93.
69. Dimick JB, Ryan AM. Methods for evaluating changes in health care policy: the difference-in-differences approach. *JAMA*. 2014;312(22):2401–2.
70. Ryan AM, Burgess JF Jr., Dimick JB. Why we should not be indifferent to specification choices for difference-in-differences. *Health Serv Res*. 2015;50(4):1211–35.
71. Luque-Fernandez MA, Schomaker M, Rachet B, Schnitzer ME. Targeted maximum likelihood estimation for a binary treatment: a tutorial. *Stat Med*. 2018;37(16):2530–46.
72. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *J Clin Epidemiol*. 2008;61(4):344–9.
73. Vandenbroucke JP, von Elm E, Altman DG, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *Int J Surg*. 2014;12(12):1500–24.
74. Pouwels KB, Widyakusuma NN, Groenwold RH, Hak E. Quality of reporting of confounding remained suboptimal after the STROBE guideline. *J Clin Epidemiol*. 2016;69:217–24.
75. Turner L, Shamseer L, Altman DG, et al. Consolidated standards of reporting trials (CONSORT) and the completeness of reporting of randomised controlled trials (RCTs) published in medical journals. *Cochrane Database Syst Rev*. 2012(11).
76. Smidt N, Rutjes AWS, van der Windt DAWM, et al. The quality of diagnostic accuracy studies since the STARD statement. Has it improved? *Neurology*. 2006;67(5):792–7.
77. Professional Society for Health Economics and Outcomes Research. Consolidated Health Economic Evaluation Reporting Standards (CHEERS) II 2020 [2020-03-09]. Available from: https://www.ispor.org/member-groups/task-forces/consolidated-health-economic-evaluation-reporting-standards-(cheers)-2.