ScienceDirect

Contents lists available at sciencedirect.com
Journal homepage: www.elsevier.com/locate/jval

Systematic Literature Review

Assessment of Studies Evaluating Incremental Costs, Effectiveness, or Cost-Effectiveness of Systemic Therapies in Breast Cancer Based on Claims Data: A Systematic Review

Marianne Luyendijk, MSc,* Robin W.M. Vernooij, PhD, Hedwig M. Blommestein, PhD, Sabine Siesling, PhD, Carin A. Uyl-de Groot, PhD

ABSTRACT

Objectives: Large secondary databases, such as those containing insurance claims data, are increasingly being used to compare the effects and costs of treatments in routine clinical practice. Despite their appeal, however, caution must be exercised when using these data. In this study, we aimed to identify and assess the methodological quality of studies that used claims data to compare the effectiveness, costs, or cost-effectiveness of systemic therapies for breast cancer.

Methods: We searched Embase, the Cochrane Library, Medline, Web of Science, and Google Scholar for English-language publications and assessed the methodological quality using the Good Research for Comparative Effectiveness principles. This study was registered with the International Prospective Register of Systematic Reviews (PROSPERO) under number CRD42018103992.

Results: We identified 1251 articles, of which 106 met the inclusion criteria. Most studies were conducted in the United States (74%) and Taiwan (9%) and were based on claims data sets (35%) or claims data linked to cancer registries (58%). Furthermore, most included large samples (mean 17 130 patients) and elderly patients, and they covered various outcomes (eg, survival, adverse events, resource use, and costs). Key methodological shortcomings were the lack of information on relevant confounders, the risk of immortal time bias, and the lack of information on the validity of outcomes. Only a few studies performed sensitivity analyses.

Conclusions: Many comparative studies of cost, effectiveness, and cost-effectiveness have been published in recent decades based on claims data, and the number of publications has increased over time. Despite the availability of guidelines to improve quality, methodological issues persist and are often inappropriately addressed or reported.

Keywords: administrative data, claims data, secondary data.

VALUE HEALTH. 2020; 23(11):1497–1508

Introduction

Data from randomized controlled trials (RCTs) are typically used to show the comparative effectiveness of treatment options and are often used in cost-effectiveness analyses to support drug reimbursement decisions.[1] However, although RCTs provide a gold standard methodology for assessing efficacy and safety, it is increasingly recognized that they fail to reflect either the true effectiveness or, if measured, the true costs of a treatment in daily practice because they use highly selected cohorts in controlled conditions.[2-4] Moreover, the primary endpoints of many phase III RCTs are only intermediate outcomes, such as relapse-free or progression-free survival, whereas final outcomes are more relevant to patients, such as survival and quality of life.[5-7] To improve our understanding of treatment effects and costs in routine clinical practice, healthcare decision makers are therefore increasingly prioritizing real-world data (RWD) collected from sources other than traditional RCTs.[8-10]

The digital era has resulted in a proliferation of RWD.[11] Large electronic databases, rich in longitudinal patient information for large cohorts, are routinely being generated as a byproduct of clinical care and financial transactions.[12] Typical examples are electronic health records and claims and billing databases,[13] and although the information they contain is not collected or stored for research purposes, it can be used to assess the costs, effectiveness, and cost-effectiveness of drugs used in routine clinical practice.[12,14] Billing and claims data, in particular, are an appealing source to researchers because they are relatively

* Address correspondence to: Marianne Luyendijk, MSc, Department of Research and Development, IKNL & ESHPM, Erasmus University, Postbus 1738, 3000 DR (Room J8.09) Rotterdam, The Netherlands. Email: m.luyendijk@iknl.nl

inexpensive (given the large sample sizes), easily accessible, and structured with codes.[14-16] Consequently, such data sets have been used not only to examine care patterns (eg, guideline adherence or regional variation) and costs but also to evaluate treatment effectiveness.[17-20]

Despite the clear benefits of claims data, several concerns must be addressed when using them for comparative research. For instance, the data may lack important clinical information (eg, diagnosis and outcomes), we may be uncertain of the accuracy of the codes, and we must acknowledge that the data may only reflect a select patient population (eg, the insured).[12] The absence of treatment randomization is another important concern inherent to all observational data.[21,22] These issues threaten the validity of study results and the usefulness of those results for decision makers. Nevertheless, many issues can be addressed through proper research design, appropriate analysis, and standardized reporting.[9] To guide comparative research based on non-RCT data (including claims data) and to assist decision makers in judging the validity of such studies, several good research practice guidelines have been developed. These include the Good Research for Comparative Effectiveness (GRACE) initiative, the ISPOR series on Good Practice for Comparative Effectiveness Research, and a technical support document by the National Institute for Health and Care Excellence.[12,23-26] The extent to which these recommendations are implemented by researchers is currently unclear.

In the present study, we aimed to describe the quantity and to assess the quality of published (cost-)effectiveness research of systemic therapies for breast cancer based on claims data. We focused on breast cancer because it is one of the most prevalent cancers worldwide, placing significant health and financial burdens on society,[27-29] and because the number of innovative therapies for this disease is increasing.[30,31]

Methods

Search Strategy

A systematic review was conducted to obtain an overview of all (cost-)effectiveness studies of systemic therapies for the treatment of breast cancer based on claims data. We primarily searched the Embase, Cochrane Library, Medline, Web of Science, and Google Scholar databases; however, this was supplemented by screening the reference lists of studies deemed relevant based on full-text reviews of the introductions and discussions of selected articles. A detailed list of the keywords used for each database can be found in Appendix File 1 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2020.05.008. The searches focused on titles and abstracts, had no time restrictions, and were restricted to articles published in English. The main database search was conducted on July 4, 2018. The study protocol was registered with the International Prospective Register of Systematic Reviews before conducting the review (number CRD42018103992).

Study Selection

The full inclusion and exclusion criteria are detailed in Appendix File 2 in Supplemental Materials. Broadly, we included studies based on the PICO framework:[32] (1) patients were diagnosed with invasive/metastatic breast cancer; (2) the analysis compared interventions (or compared with none) based on patient-level claims data (including those linked to other sources); (3) interventions included systemic anticancer treatments, with or without radiotherapy or surgery, and with no restriction on comparators; and (4) outcomes were either costs (or resource use), effects (clinical outcomes, adverse events, including cancer recurrence and the development of other disease in later life, treatment switching, or patient-reported outcomes), or both (ie, cost-effectiveness). Titles and abstracts were screened for inclusion by 2 independent reviewers (ML and RV) before they reviewed the full texts of articles to identify those eligible for data extraction. At both stages of article selection, disagreements were resolved by discussion, deferring to a third reviewer (SS or HMB) to make the final decision when no agreement could be reached.

Data Extraction

Data were extracted from each study using a form designed by ML and pilot tested by ML and RV. The following data were collected: author, year of publication, country, time horizon, study design, database, sample size, patient selection method, patient characteristics, treatment type, comparison type, outcome type, outcome measures, and statistical methods.

Quality Assessment

All included studies were assessed for methodological quality by 2 independent researchers (ML and RV), using the validated GRACE checklist. This checklist was specifically developed to evaluate the quality of comparative effectiveness research based on observational data.[26] It comprises 6 items concerning data quality (eg, availability of information in the data set) and 5 items concerning the methodology (eg, study design and analysis). Additional criteria were also used to improve consistency among the researchers and across the assessment of the included publications. These criteria were based on previous literature,[33,34] pragmatic literature searches, and the researchers' judgments. The full checklist, including descriptions of the main and additional criteria, can be found in Table 1 and Appendix File 3 in Supplemental Materials (found at https://doi.org/10.1016/j.jval.2020.05.008). We planned to resolve disagreements in the quality assessment through discussion until consensus was reached.

Data Analysis

The characteristics of the selected studies are presented as numbers and percentages for categorical variables and as means and standard deviations (minimum to maximum) for continuous variables. For the quality assessment, we estimated the proportion of studies that fulfilled each criterion and compared the GRACE scores before and after 2010. This cutoff was chosen because good research guidelines had been published at the end of 2009.[12,24,35,36]

Results

Descriptive Statistics

We identified 1251 unique studies, and we excluded 1047 based on title and abstract review and 98 based on full-text review. Thus 106 studies were included for data extraction and quality assessment (Figure 1; Appendix Files 4 and 5 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2020.05.008).

Study Details

A full list of the included studies and their details is presented in Appendix File 6 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2020.05.008. In the following text, we describe specific features of the included studies in more detail.

Table 1. GRACE checklist and additional criteria.

	Description	Answer options GRACE[59]	Additional criteria	Source (additional criteria)
D1	Were treatment and important details of treatment exposure adequately recorded for the purpose of the study?	**YES:** Reasonably necessary information to determine treatment or intervention was adequately recorded for study purposes (eg, for drugs, sufficient detail on dose, days supplied, route, or other important data).	**YES:** If there is some information on dose/duration (eg, no. of treatment lines).	33
		NO: Data source clearly deficient or there is not enough information in article.	**NO:** No information on dose/ duration.	
D2	Were the primary outcomes adequately recorded for the purpose of the study (eg, available in sufficient detail through data source[s])?	**YES:** Information to ascertain outcomes were adequately recorded in the data source (eg, if clinical outcomes were ascertained using ICD-9 diagnosis code[s] in an administrative database, the level of sensitivity and specificity captured by the code[s] was sufficient for assessing the outcome of interest.)	**YES**: 1) The outcome(s) based on algorithm/codes and the sensitivity, specificity, or PVV is reported in the article or in the article referred to. 2) Mortality outcomes are based on the cancer registry and/or death registry. 3) Other outcomes obtained through medical chart review.	
		NO: Data source clearly deficient (eg, the code(s) captured a range of conditions that was too broad or narrow, and supplementary information such as that from medical charts was not available), or not enough information in article.	**NO:** One or more of the outcomes used for the analyses is based on an algorithm or codes AND the algorithm/code is not validated and/or sensitivity, specificity, or PVV are not reported in the paper or reference.	
D3	Was the primary clinical outcome(s) measured objectively rather than subject to clinical judgment (eg, opinion about whether the patient's condition has improved)?	**YES:** clinical outcomes were measured objectively (eg, hospitalization, mortality). **N/A:** primary outcome not clinical (eg, PROs).	N/A.	
		NO: eg, clinical opinion about whether patient's condition improved, or not enough information in article.	N/A.	
D4	Were primary outcomes validated, adjudicated, or otherwise known to be valid in a similar population?	**YES:** Outcomes were validated, adjudicated, or based on medical chart abstractions with clear definitions (eg, a validated instrument was used to assess patient reported outcomes, such as the SF-12 Health Survey); a clinical diagnosis via ICD-9 code was used, with formal medical record adjudication by committee to confirm the diagnosis or other procedures to achieve reasonable sensitivity and specificity; billing data were used to assess health resource utilization, etc.	**Yes**: 1) The outcome(s) based on algorithm/codes and the sensitivity, specificity, or PVV is reported in the article or in the article referred to. 2) Mortality outcomes based on the cancer registry and/or death registry. 3) Other outcomes obtained through medical chart.	
		NO: No, or not enough information in article.	**NO:** One or more of the outcomes used for the analyses is based on an algorithm or codes AND the algorithm/code is not validated and/or sensitivity, specificity, or PVV are not reported in the paper or reference.	
D5	Was the primary outcome(s) measured or identified in an equivalent manner between the treatment/intervention group and the comparison group(s)?	**YES.**	N/A	
		NO, or not enough information in article.	N/A	

continued on next page

Table 1. Continued

	Description	Answer options GRACE[59]	Additional criteria	Source (additional criteria)
D6	Were important covariates that may be known confounders or effect modifiers available and recorded?	**YES:** Most if not all important known confounders and effect modifiers were available and recorded (eg, measures of medication dose and duration).	**YES:** A list of important confounders/covariates was determined with a pragmatic literature search per outcome and disease stage (see Appendix File 3 in Supplemental Materials). This item was judged to be sufficient (YES) if the confounders/covariates in the list were available in the data set of the study.	Appendix File 3 in Supplemental Materials
		NO: At least one important known confounder or effect modifier not available and recorded (as noted by authors or as determined by user's clinical knowledge), or not enough information in article.	**NO:** If one or more of the specified confounders/covariates were missing.	Appendix File 3 in Supplemental Materials
M1	Was the study (or analysis) population restricted to new initiators of treatment or those starting a new course of treatment?	**YES**: Only new initiators of the treatment of interest were included in the cohort, or for surgical procedures and devices, only patients who never had the treatment before the start of study follow-up were included.	**YES:** If efforts were made to exclude patients who could have had the treatment before. Or if it is very unlikely that patients had the treatment before (eg, first treatment after first primary diagnosis of breast cancer).	
		NO: or not enough information in article.	**NO:** If it is possible that patients had the treatment before and no efforts were made to check if patients had the treatment before diagnosis (eg, enrollment in insurance at least a certain period before diagnosis) and/or to exclude patients who could have had the treatment before (eg, patients with other cancer before breast cancer or recurrent disease).	
M2	If one or more comparison groups were used, were they concurrent comparators? If not, did the authors justify the use of historic comparisons group(s)?	**YES:** Data were collected during the same time period as the treatment group ("concurrent") or historic comparators were used with reasonable justification (eg, when it is impossible for researchers to identify current users of older treatments or when a concurrent comparison group is not valid—ie, uptake of new product is so rapid that concurrent comparators differ greatly on factors related to the outcome).	**YES**: The timeframe of patients selection was $\leq$ 3 years (eg, patients diagnosed with breast cancer between 2002 and 2004) or the timeframe was >3 years and it was evident that treatments were provided during the same time period.	
		NO: Historic comparators were used without being scientifically justifiable, or there was not enough information in article.	**NO:** The timeframe of patient selection was > 3 years and it was unclear when treatments (ie, intervention and comparator[s]) were provided.	
M3	Were important covariates and confounding and effect modifying variables considered in the design and/or analysis?	**YES:** Most if not all important covariates that would be likely to change the effect estimate substantially were accounted for, (eg, measures of medication dose and duration).	**YES:** A list of important confounders/covariates was determined with a pragmatic literature search per outcome and disease stage (Appendix File 3 in Supplemental Materials). This item was judged to be sufficient (YES) if the confounders/covariates in the list were considered in the analysis.	Appendix File 3 in Supplemental Materials

continued on next page

Table 1. Continued

	Description	Answer options GRACE[59]	Additional criteria	Source (additional criteria)
		NO: Some important covariates were available for analysis but not analyzed appropriately, or at least one important covariate was not measured, or there was not enough information in article.	**NO:** If one or more of the specified confounders/covariates were not considered in the analysis.	Appendix File 3 in Supplemental Materials
M4	Is the classification of exposed and unexposed person-time free of ITB?	**YES**	**YES**: If the study is at low risk of ITB: 1) Time-dependent covariates analysis was used. 2) Landmark analyses or restriction in selection of patient populations (eg, patients had to be alive at a certain period). 3) If it was highly unlikely or impossible for an outcome to occur before the start of treatment (eg, recurrence/mortality before the start of adjuvant chemotherapy in early stage breast cancer). 4) Comparison of two or more treatments without a comparison to no treatment.	34
		NO: Or not enough information in this article.	**NO:** At high risk of ITB.	
M5	Were any meaningful analyses conducted to test key assumptions on which primary results are based?	**Yes:** And primary results did not change substantially. **Yes:** But primary results changed substantially.	N/A.	
		None reported, or not enough information in article.	N/A.	

GRACE indicates Good Research for Comparative Effectiveness; ICD, International Classification of Diseases; ITB, immortal time bias; PROs, patient-reported outcome measures; PVV, positive predictive value; SF-12, Short Form.

Publication year, country, database, design

The earliest study in this review was published in 2002, and the number of studies showed a trend to increase over time (Fig. 2a). Most studies originated from the United States (78 studies; 74%), Taiwan (10 studies; 9%), and Canada (7 studies; 7%), with the remainder from other countries (11 studies; 10%) (Fig. 2b). About one-third (37 studies; 35%) used claims data not linked to a database with patient and/or clinical characteristics, but the remainder linked claims data with either cancer registry data (61 studies; 58%) or other sources (eg, RCTs; 8 studies; 7%) (Table 2). The Surveillance Epidemiology and End Results–Medicare linkage database was used most frequently (41 studies; 39%), followed by the Truven MarketScan database (claims data only; 12 studies; 11%) and the Taiwan National Health Insurance Research Database (claims data only; 10 studies; 9%). Finally, 83 (78%) of the studies compared only the effectiveness of treatments, 12 studies (11%) compared only the costs of different therapies, 9 (9%) estimated both the effects and incremental costs of therapies, and 2 (2%) performed full cost-effectiveness analyses, reporting the costs, effects, and incremental cost-effectiveness ratios (see Table 2).

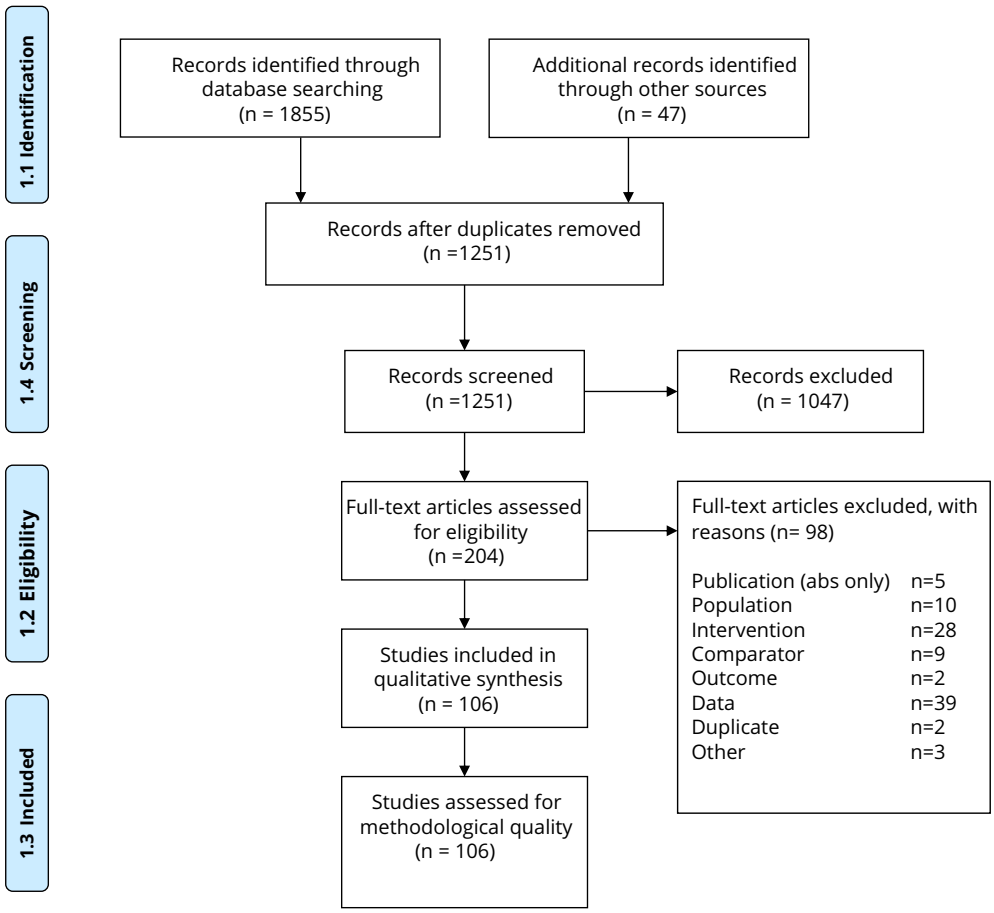
Patient populations, sample size, cohort selection

Study populations comprised patients with locoregional disease (43 studies; 41%), metastatic disease (27 studies; 25%), any stage, (21 studies; 20%), or unclassified stage (15 studies; 14%). They also focused on patients aged ≥65 years (43 studies; 41%), adults aged about ≥18 years (29; 27%), or patients with no age limit specified (22 studies; 21%). If specified, patients were rarely younger than 18 or 65 years. Sample sizes ranged from 175 to 190 620 (mean: 17 130; median: 6433). Samples for the linked data sets were mostly selected based on cancer registry data (52 studies; 49%) or a combination of an algorithm and cancer registry data (6; 6%), whereas claims-only studies were selected based on algorithms (33; 31%) or single claim codes (7; 7%). Unfortunately, most of the algorithms/codes were not validated (see Table 2).

Treatment and comparator

Chemotherapy (CT) was evaluated in 46 studies (43%). These either assessed all agents, a specific subclass, or a prespecified regimen (eg, doxorubicin, cyclophosphamide, and taxane). Hormonal therapy (HT) was assessed in 27 studies (26%), typically evaluating the risk of developing health problems in the future (eg, cardiovascular problems and diabetes) or the effect of treatment adherence on survival and related outcomes. Another 13 studies (12%) evaluated targeted therapy (TT), 3 (3%) compared CT with HT, 12 (11%) compared CT with TT, and some did not specify the treatment (eg, "any" systemic therapy). The comparator varied among studies. Over one-third (41 studies; 39%) included no therapy as a comparator (eg, HT vs no HT), and about one-third (36 studies; 34%) compared 2 or more therapies or regimes (eg, some compared HT and CT whereas others compared multiple different CT regimens). Some studies compared the effects of different times to initiation of therapy (13 studies; 12%) and the effects of adherence and nonadherence (12 studies; 11%).

Figure 1. PRISMA flowchart for the inclusion and exclusion of publications.



PRISMA indicates Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

Outcomes

Multiple outcomes were evaluated in most studies (see Table 2), mainly related to survival (ie, overall or breast cancer specific survival), adverse events, and recurrence. Other frequently studied outcomes included healthcare visits and costs. Of the studies that assessed costs, 5 measured cancer-related costs, 16 measured total costs (ie, costs related to cancer care and costs of unrelated conditions), and 2 measured both types.

Figure 2. Publications per year and countries of the studies (N = 106).

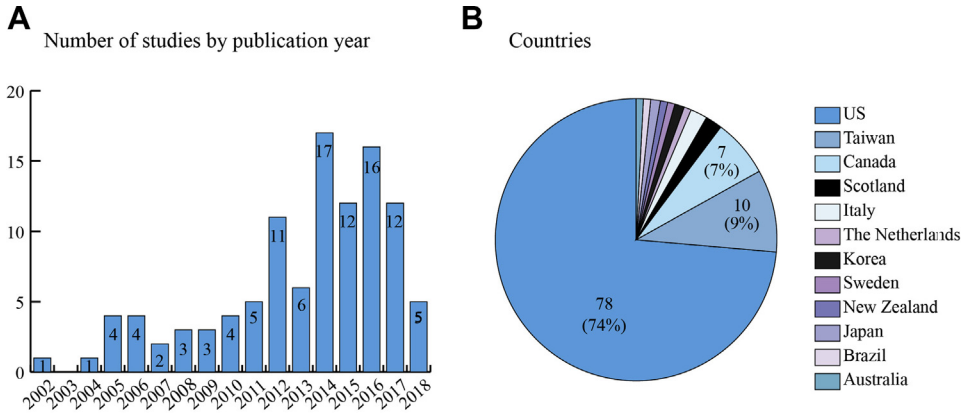


Table 2. Key characteristics of the 106 included studies.

	N	%
Total number of studies	106	100%
Types of databases		
Claims data	37	35%
Claims data linked to cancer registry data	61	58%
Claims data linked to other sources	8	7%
Study design		
Comparative effectiveness	83	78%
Costs comparison	12	11%
Comparison of effects and costs separately	9	9%
Cost-effectiveness	2	2%
Patient population		
Disease		
Locoregional	43	41%
Metastatic	27	25%
Both	21	20%
Not specified/other*	15	14%
Age group		
About ≥18 years†	29	27%
About ≥65 years†	43	41%
All patients‡	22	21%
Other§	12	11%

	Mean (SD)	Median (min/max)	N
Data			
Sample size	17 130 *(26,438)*	6433 *(175/190,620)*	106
Study period in years‖	7.8 *(4.7)*	7.0 (0/lifetime¶)	106
Median follow-up in months#	40 *(22)*	n/a	39
Treatment and comparator			
Comparison of**:			
a) ≥1 therapy vs no therapy	41		39%
b) ≥2 therapies	36		34%
c) Different timing††	13		12%
d) Adherence vs non-adherence‡‡	12		11%
e) Other	4		4%
Treatment type			
Chemotherapy	46		43%
Hormone therapy	27		26%
Targeted therapy	13		12%
Chemotherapy & hormone therapy	3		3%
Chemotherapy & targeted therapy	12		11%
Not specified	3		3%
All	2		2%
Outcomes	**N = 173**§§		
Overall survival	37		21%
(Breast) cancer specific survival	16		9%
Recurrence or other cancer	16		9%
Treatment switching	9		5%
Adverse events‖‖	48		28%
Healthcare costs (total costs/cancer related costs /both)	23 (n = 16/n = 5/n = 2)		13%
Healthcare visits	18		11%
Other¶¶	7		4%
Cohort selection			
Cancer registry	52		49%
Cancer registry + algorithm	6		6%
Algorithm – adapted from literature	11		10%
Algorithm – validated	2		2%
Algorithm – not validated	20		19%
Single code with BC diagnosis	7		7%
Other	8		8%
Statistical methods for selection bias	**N = 140**##		
Stratification	19		14%
Regression analyses (multivariate)	82		58%

continued on next page

Table 2. Continued

	Mean (SD)	Median (min/max)	N
Linear regression	3		Total >82; multiple methods were used in some studies
Log/logit regression	14		
Poisson regression	2		
Negative binominal	2		
Tobit regression	1		
Generalized linear models***	11		
Cox regression	46		
Competing risk regression	1		
Cox regression with time-dependent confounders/covariates	16		
Propensity score methods	27		19%
Propensity score matching	12		Total >27; multiple methods were used in some studies
Propensity score weighting	6		
Propensity score stratification	4		
Propensity score covariate adjustment	11		
Doubly robust methods†††	2		
Not specified	2		
Instrumental variables	0		0%
None	11		8%
Other‡‡‡	1		1%

BC indicates breast cancer; max, maximum; min, minimum; SD, standard deviation.
*Other is for stages 1 and 2 or stages 2 to 4.
†The category "about ≥18 years" includes studies that defined the age as approximately 18 without an upper age limit, the category "≥65" defined the age as approximately 65.
‡In these studies age was not specifically specified (eg, patients with invasive BC), so we assumed no age limits were applied.
§Other includes >50/<65/>80, postmenopausal, and so on.
‖The study period includes the period for selection of patients + follow-up.
¶Lifetime was assumed to be 34 years as in this study patients of ≥65 years were included and we assumed that patients were max 99 years old (99 − 65 = 34) and 0 years of follow-up were for a cross-sectional study.
#Median follow-up was reported by 38 studies; other studies reported mean follow-up/person years or did not report follow-up duration.
**Some studies made multiple comparisons (eg, comparison of treatment vs not treatment and comparison of timing of the treated); we included only the main comparison.
††Different timing includes, for instance, a delay in treatment initiation, with treatment restarting after discontinuation.
‡‡Most studies that compared adherence with nonadherence also compared persistence with nonpersistence.
§§Many studies evaluated multiple outcomes; the number of outcomes therefore sum up to more than 106.
‖‖Short-term adverse events (AEs), such as hospitalizations for neutropenia, fever, and thrombocytopenia, and long-term AEs (eg, development of cardiovascular events/diabetes/depressive disorders).
¶¶Other outcomes (eg, out of pocket payments, incremental cost-effectiveness ratio, noncancer survival)
##Some studies used more statistical methods; thus it does not add up to 106.
***This includes generalized linear models and extended estimation equations.
†††Combine the IPTW and regression model.
‡‡‡Randomization.

Statistical methods to control for selection bias

Most studies used multiple methods to control for selection bias, including regression analyses and propensity score matching, but none used an instrumental variable method. Only 11 studies did not attempt to control for confounding, but most of these did not intend to draw inferences about the relative effectiveness/costs or were pilot studies for which sample sizes were too small to control for confounders (see Table 2).

Quality Assessment GRACE Checklist

The results of the quality assessment are summarized in Table 3 and Appendix File 7 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2020.05.008. Although agreement was reached on all items, a couple of points are noteworthy. First, we had some difficulties reaching consensus on the issue of immortal time bias (ITB), so we asked a third reviewer to check 25 studies that we had initially disagreed on; however, this did not change our original conclusions. Second, we also had extensive discussions about the issue of relevant confounders, although again, we ultimately agreed on all items.

Only 17 studies (16%) reported information on the dose or duration of the study treatment (item D1). Fewer than half of the studies reported the accuracy of outcomes based on the International Classification of Disease/claim code(s) or algorithms. Given that many studies used adverse events or cancer recurrence as outcomes, GRACE items D2 and D4, which evaluate the recording and validity of primary outcomes, were not fulfilled by 63 (59%) and 62 (58%) studies, respectively. Nevertheless, the primary outcomes of most studies were based on either International Classification of Disease/claim codes or death registry data, resulting in 104 studies (98%) using objective outcomes (item D3). Given that the same claim codes/mortality data were used for the intervention and comparison group, 100% of the studies fulfilled the criteria for item D5. Many data sets lacked information on relevant confounders for the outcome, with 47 studies (44%) each meeting the criteria for items D6 and M3. For example, information on major risk factors (eg, lifestyle indicators for cardiovascular diseases), disease severity (eg, stage), performance status, and HER2 status were frequently unavailable in the data sets (see Appendix File 3 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2020.05.008 for a full list of confounders deemed relevant per outcome).

In most studies, patients were selected based on specified criteria, such as the type of breast cancer and the period of diagnosis. Although most described a long timeframe for selection (eg, diagnosed between 1998 and 2008), they often failed to describe when the treatments under comparison were given. As a result, item M2 was deemed to be sufficient in only 26 studies (25%)

Table 3. Quality assessment using the GRACE checklist.

Item	Question	Total	Preguidelines (2002-2009)	Postguidelines (2010-2018)
		N = 106	N = 18	N = 88
D1	Were treatment (details) recorded adequately?	17 (16%)	3 (17%)	14 (16%)
D2	Were outcomes recorded adequately?	43 (41%)	7 (39%)	36 (41%)
D3	Was the primary clinical outcome(s) measured objectively?	104 (98%)	18 (100%)	86 (98%)
D4	Was the primary outcome validated?	44 (42%)	7 (39%)	37 (42%)
D5	Were primary outcomes measured/identified in an equivalent manner between the treatment and comparator groups?	106 (100%)	18 (100%)	88 (100%)
D6	Were important covariates recorded?	47 (44%)	11 (61%)	36 (41%)
M1	Was the study restricted to new initiators of treatment?	45 (42%)	8 (44%)	37 (42%)
M2	Were comparison groups concurrent comparators?	26 (25%)	3 (17%)	23 (26%)
M3	Were important confounders considered?	47 (44%)	11 (61%)	36 (41%)
M4	Is the analysis free of "immortal time bias?"	82 (77%)	15 (83%)	67 (76%)
M5	Were any sensitivity analyses performed?	37 (35%)	5 (28%)	32 (36%)

Note. Data are presented as n (%) of studies that fulfilled each criterion.
GRACE indicates Good Research for Comparative Effectiveness.

because we could not be certain that data were collected during the same period in 80 studies (75%). Authors sometimes required the inclusion of information about patients and treatments before the study follow-up period, such as requiring patients to be enrolled in an insurance program for at least 12 months before diagnosis so that the authors could select new treatment initiators for analysis. Only 45 studies (42%) therefore met the criteria for item M1.

Item M4 addressed the issue of ITB. Immortal time refers to a period during cohort follow-up in which the event of interest could not have occurred, with bias arising when this is not correctly dealt with in the study design or analysis.[37,38] The risk of ITB was high in 24 studies (23%) in our review; indeed, these made no attempts to deal with the issue in their design or analysis, despite the high likelihood that the event occurred before the exposure definition was fulfilled. Most of these studies evaluated treatment adherence and persistence or compared exposed and non-exposed patients over long periods within the definition of exposure (eg, ever vs never users).

Finally, only 37 studies (35%) reported a sensitivity analysis (item M5). Nevertheless, unexpected differences were observed from before to after the good practice guidelines were published (see Table 3). Notably, after their publication, fewer studies included relevant confounders/covariates (items D6 and M3) and more studies appropriately reported sensitivity analyses (item M5).

Discussion

We conducted a systematic review of studies assessing the incremental costs, effectiveness, or cost-effectiveness of systemic therapies for breast cancer based on claims data and rated the methodological quality of 106 included studies using the GRACE checklist. The earliest study was published in 2002, most were published in the United States and Taiwan, and there was a clear increase in number over time. Nevertheless, many studies had methodological shortcomings, and it was notable that the quality of studies did not improve after the publication of good practice guidelines.

The observed trend in the number of studies is not surprising given the proliferation of RWD and the growing interest in its use to improve clinical and regulatory decision making for oncology.[12,39,40] RWD is believed to have the potential to complement evidence from RCTs thanks to certain well-known benefits, of which several were evident in the studies included in this review. First, we should note that many studies were based on the Surveillance Epidemiology and End Results Medicare–linked database in the United States, which has opened more opportunities to conduct studies using RWD. Second, the sample sizes were generally large, ranging from 175 to 190 620 (even covering entire countries, such as Taiwan[41]), which contrasts dramatically with the numbers in clinical trials (ie, only 3% to 5% of all patients with cancer).[40] Third, about 40% of the studies included older patients (typically ≥65 years) and some even included the very old (>80 years), groups that are known to be underrepresented in RCTs.[42] Fourth, diverse outcomes were studied in the claims-based studies, including healthcare resource utilization, costs, survival, and late or rare adverse events. These outcomes can be difficult to study in RCTs because of the controlled treatment settings, small sample sizes, and limited follow-up durations.[7,43,44] Finally, claims studies allow for evaluations of different therapies that are not typically compared in head-to-head clinical trials; for example, combinations of doxorubicin, cyclophosphamide, paclitaxel, and trastuzumab were compared with docetaxel, carboplatin, and trastuzumab.[20,45]

Despite the advantages of using RWD, our review also showed that important methodological requirements were not always met in published studies. One frequently observed issue, which is particularly relevant for comparative studies, was the lack of information about confounding variables. Most studies did link claims and cancer registry data, thereby including tumor and patient information (eg, cancer stage), and many used proxies of health status (eg, comorbidity index and performance status).[46,47] Important confounders for the outcome of interest, however, were frequently lacking. In several studies that examined the effect of systemic treatments on cardiovascular problems, information on risk factors such as obesity or smoking status were not available in either the claims or the registry data.[48] This absence of prognostic data is an established issue of RWD, especially in retrospective

databases where researchers do not control the data collection.[49] In these settings, failure to include all relevant confounders and to exclude all irrelevant confounders can lead to biased estimates of treatment effects.[24]

Good practice guidelines recommend that all potential confounding factors be identified and that researchers preferably make use of directed acyclic graphs (DAGs) to visualize the relationships among variables.[24] Such approaches make it possible to recognize relevant and irrelevant confounders, and they can guide researchers when interpreting model results.[24] In our review, very few studies provided clear rationales for the choice of potential confounding variables, and none used DAGs. Given the trend toward using large longitudinal secondary databases for comparative research, we believe that it will become ever more critical to discuss and defend the reasons for including or excluding confounders. This transparency makes it easier for decision makers to judge the reliability of the observed effect.

Another quality concern that was quite common in the claims-based studies of this review was the high risk of ITB. This phenomenon can arise when there is a time period during cohort follow-up in which the event of interest cannot occur. Not appropriately accounting for this can produce misleading results (ie, an overestimate of treatment effects).[37,38] Previous systematic reviews have assessed the prevalence of ITB in observational studies published in high-quality medical journals.[34,50] Consistent with our research, they also found it to be surprisingly common.[34,38,50,51] Because ITB can be difficult for readers to recognize, we believe that researchers who use claims data should explicitly specify how they dealt with ITB (eg, Chien et al),[41] in line with the RECORD-PE checklist (Reporting of Studies Conducted using Observational Routinely Collected Health Data Statement for Pharmacoepidemiology).[52]

We also found that outcomes based on claim codes or algorithms were often not validated and that performance characteristics (ie, predictive values, sensitivity, specificity) were almost never provided. Validating outcomes is important because codes in claims data are not always accurate, and using inaccurate codes can lead to misclassification and loss of internal validity if the occurrence is not at random (ie, classification bias).[24] Such misclassification can vary by disease state, patient population, data source, and code.[24,53] Algorithms for cancer recurrence, an outcome frequently used in many studies in our review, have been shown to vary widely depending on the code and type of cancer. For example, the positive predictive values for breast, lung, and prostate cancers have been reported to be 30% to 87%, 72% to 94%, and <30%, respectively.[54] This supports the argument that authors should report their algorithm's performance to facilitate the accurate interpretation of their results.

Other issues were identified with the GRACE checklist. Few studies performed sensitivity analyses to explore how much of the estimated effect depended on underlying assumptions. This is an important omission because many assumptions are made about the patient population, the exposure, and the outcomes in claims-based studies, and because sensitivity analyses can provide insights into the extent to which study findings are dependent on them. Few studies also reported details of the doses or durations of the treatments under study, possibly because that information was unavailable or unreliable in their claims datasets.[55] However, such information is particularly relevant when considering toxicity outcomes. Finally, it was often unclear whether the compared treatments were prescribed at the same or at different times during the study periods, preventing the reader from assessing the impact of change in the standard of care over time.

When using a checklist to grade the quality of studies, assessors must perform valid and reliable interpretations. We ensured validity by using the validated GRACE checklist for evaluating observation studies[26] and ensured reliability by using 2 independent assessors and a checklist clarification. Nevertheless, we still believe that aspects other than those listed in the GRACE checklist are relevant when evaluating studies based on claims data. For instance, the checklist does not cover either the selection of participants or the measurement of treatment exposure, and both are often identified by algorithms or codes and can result in misclassification. We found that few studies relying on claims databases alone used validated codes or algorithms to select samples. A similar finding was reported by Schulman et al (2014), who subsequently developed a checklist for selecting study cohorts in oncology research.[56] This checklist appears to complement the GRACE checklist, and it could be suitable to combine these in future research.[57] Another highly relevant consideration for claims-based studies is how results are interpreted in the context of other literature on the topic of interest. This is because confounding can never be totally excluded in the absence of randomization, which necessitates proper interpretation to increase confidence in the reliability of the direction and magnitude of observed findings.[12] This is not included in the GRACE checklist, but it was raised in the discussions of many studies in this review.

We performed a comprehensive bibliographic database search using various combinations of MeSH/Emtree terms, free text terms, and synonyms to identify the studies included in this review. Nevertheless, it was difficult to find all the studies that used claims data because authors often used the database name rather than a general term for the data set, such as "claims data" or "administrative data." Moreover, MeSH/Emtree terms for claims data were only recently added to the bibliographic databases we used. To reduce the impact of this limitation, we also manually screened the references of all included publications. It should be noted that the large number of studies from the United States could result from the bias of including the MeSH/Emtree terms "Medicare" and "Medicaid" (US insurance programs). We nevertheless believe that this will have had little effect on our findings and recommendations because only minor differences in study quality were identified between those conducted in the United States and in other countries (Appendix File 8 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2020.05.008).

Finally, few studies specified whether they were exploratory or hypothesis testing in nature, and moreover, there was a conspicuous lack of information to help reviewers consistently distinguish between these 2 types. According to Berger et al (2017), authors should specify this design element before conducting studies of treatment effectiveness.[58] This is because, whereas hypothesis testing studies seek to support decision making, exploratory studies seek to generate hypotheses for further research.[58] Therefore the requirements and practice recommendations for studies testing hypotheses of treatment effectiveness must be stricter. Providing greater clarity about the study objective in the aims, and specifically about whether a study is exploratory or hypothesis testing in nature, could help readers and decision makers judge whether a given study has sufficient quality for their needs.

Conclusion

Many comparative (cost)-effectiveness studies based on claims data have been published in recent decades, and the number of publications has clearly increased over this time. Our review highlights that methodological issues are frequently not addressed or reported appropriately despite the availability of good practice and reporting guidelines. Adherence to these guidelines must improve before the promise of claims data to

increase insights into the effectiveness of cancer treatments can be fulfilled.

Supplemental Material

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jval.2020.05.008.

Article and Author Information

Accepted for Publication: May 11, 2020

Published Online: September 10, 2020

doi: https://doi.org/10.1016/j.jval.2020.05.008

Author Affiliations: Department of Research and Development, Netherlands Comprehensive Cancer Center, Utrecht, The Netherlands (Luyendijk, Vernooij, Siesling); Erasmus School of Health Policy and Management, Erasmus University, Rotterdam, The Netherlands (Luyendijk, Blommestein, Uyl-de Groot); Department of Health Technology and Services Research, University of Twente, Enschede, The Netherlands (Siesling)

Author Contributions: *Concept and design*: Luyendijk, Vernooij, Blommestein, Siesling, Uyl-de Groot
Acquisition of data: Luyendijk, Vernooij, Blommestein, Uyl-de Groot
Analysis and interpretation of data: Luyendijk, Vernooij, Blommestein, Siesling, Uyl-de Groot
Drafting of the manuscript: Luyendijk, Vernooij, Blommestein, Siesling, Uyl-de Groot
Critical revision of the paper for important intellectual content: Siesling, Uyl-de Groot
Supervision: Vernooij, Blommestein, Siesling, Uyl-de Groot

Conflict of Interest Disclosures: Dr Blommestein reported receiving grants from Celgene BV outside the submitted work. No other disclosures were reported.

Funding/Support: The study was supported by The Netherlands Comprehensive Cancer Organisation and the Erasmus School of Health Policy & Management.

Role of the Funder/Sponsor: The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Acknowledgments: We thank Gerdien de Jong (biomedical information specialist) for providing support with the search strategy of our literature review and Dr Robert Sykes for providing editorial assistance in the final drafts of this manuscript.

REFERENCES

1. Makady A, van Veelen A, Jonsson P, et al. Using real-world data in health technology assessment (HTA) practice: a comparative study of five HTA agencies. *PharmacoEconomics*. 2018;36:359–368.
2. Booth C, Tannock I. Randomised controlled trials and population-based observational research: partners in the evolution of medical evidence. *Br J Cancer*. 2014;110:551–555.
3. Marshall DA, Hux M. Design and analysis issues for economic analysis alongside clinical trials. *Med Care*. 2009;47:14–20.
4. Sherman RE, Anderson SA, Dal Pan GJ, et al. Real-world evidence - what is it and what can it tell us? *N Engl J Med*. 2016;375:2293–2297.
5. Del Paggio JC, Azariah B, Sullivan R, et al. Do contemporary randomized controlled trials meet ESMO thresholds for meaningful clinical benefit? *Ann Oncol*. 2017;28:157–162.
6. Booth CM, Cescon DW, Wang L, et al. Evolution of the randomized controlled trial in oncology over three decades. *J Clin Oncol*. 2008;26:5458.
7. Tannock IF, Amir E, Booth CM, et al. Relevance of randomised controlled trials in oncology. *Lancet Oncol*. 2016;17:560–567.
8. Moen F, Svensson J, Steen Carlsson K. Assessing the value of cancer treatments from real world data—issues, empirical examples and lessons learnt. *J of Cancer Policy*. 2017;11:32–37.
9. Garrison Jr LP, Neumann PJ, Erickson P, et al. Using real-world data for coverage and payment decisions: the ISPOR Real-World Data Task Force report. *Value Health*. 2007;10:326–335.
10. Makady A, de Boer A, Hillege H, et al. What is real-world data? A review of definitions based on literature and stakeholder interviews. *Value Health*. 2017;20:858–865.
11. Booth CM, Karim S, Mackillop WJ. Real-world data: towards achieving the achievable in cancer care. *Nat Rev Clin Oncol*. 2019;16:312–325.
12. Berger ML, Mamdani M, Atkins D, et al. Good research practices for comparative effectiveness research: defining, reporting and interpreting nonrandomized studies of treatment effects using secondary data sources: the ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report–Part I. *Value Health*. 2009;12:1044–1052.
13. Khozin S, Blumenthal GM, Pazdur R. Real-world data for clinical evidence generation in oncology. *J Natl Cancer Inst*. 2017;109(11):1–5.
14. Liang SY, Phillips KA, Wang G, et al. Tradeoffs of using administrative claims and medical records to identify the use of personalized medicine for patients with breast cancer. *Med Care*. 2011;49:1–8.
15. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J Clin Epidemiol*. 2005;58:323–337.
16. Birnbaum HG, Cremieux PY, Greenberg PE, et al. Using healthcare claims data for outcomes research and pharmacoeconomic analyses. *Pharmacoeconomics*. 1999;16:1–8.
17. Potosky AL, Riley GF, Lubitz JD, et al. Potential for cancer related health services research using a linked Medicare-tumor registry database. *Med Care*. 1993;31:732–748.
18. Ruddy KJ, Van Houten HK, Sangaralingham LR, et al. Impact of treatment regimen on acute care use during and after adjuvant chemotherapy for early-stage breast cancer. *Breast Cancer Res Treat*. 2017;164:515–525.
19. Dalal AA, Gagnon-Sanschagrin P, Burne R, et al. Dosing patterns and economic burden of palbociclib drug wastage in HR+/HER2- metastatic breast cancer. *Adv Ther*. 2018;35:768–778.
20. Reeder-Hayes KE, Meyer AM, Hinton SP, et al. Comparative toxicity and effectiveness of trastuzumab-based chemotherapy regimens in older women with early-stage breast cancer. *J Clin Oncol*. 2017;35:3298–3305.
21. West S, Ritchey M, Poole C. Validity of pharmacoepidemiologic drug and diagnosis data. In: Strom B, Kimmel S, Hennessy S, eds. *Textbook of Pharmacoepidemiology*. Oxford: Wiley Blackwell; 2013.
22. Toh SAS, Andrade SE, Raebel MA, et al. Examples of existing automated databases. In: Strom B, Kimmel S, Hennessy S, eds. *Textbook of Pharmacoepidemiology*. Oxford: Wiley Blackwell; 2013.
23. Faria R, Alava M, Manca A, et al. NICE DSU technical support document 17: the use of observational data to inform estimates of treatment effectiveness in technology appraisal: methods for comparative individual patient data 2015. Available from: http://www.nicedsu.org.uk. Accessed July 4, 2018.
24. Cox E, Martin BC, Van Staa T, et al. Good research practices for comparative effectiveness research: approaches to mitigate bias and confounding in the design of nonrandomized studies of treatment effects using secondary data sources: the International Society for Pharmacoeconomics and Outcomes Research Good Research Practices for Retrospective Database Analysis Task Force Report-Part II. *Value Health*. 2009;12:1053–1061.
25. 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 Health*. 2014;17:143–156.
26. Dreyer NA, Bryant A, Velentgas P. The GRACE checklist: a validated assessment tool for high quality observational studies of comparative effectiveness. *J Manag Care Spec Pharm*. 2016;22:1107–1113.
27. Fitzmaurice C, Akinyemiju TF, Al Lami FH, et al. Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 29 cancer groups, 1990 to 2016: a systematic analysis for the global burden of disease study. *JAMA*. 2018;4(11):1553–1568.
28. Luengo-Fernandez R, Leal J, Gray A, et al. Economic burden of cancer across the European Union: a population-based cost analysis. *Lancet Oncol*. 2013;14:1165–1174.
29. Mariotto AB, Yabroff KR, Shao Y, et al. Projections of the cost of cancer care in the United States: 2010-2020. *J Natl Cancer Inst*. 2011;103:117–128.
30. Li Z, Kang Y. Emerging therapeutic targets in metastatic progression: Aa focus on breast cancer. *Pharmacol Ther*. 2016;161:79–96.
31. QuintilesIMS. *Global Oncology Trends 2017: Advances, Complexity and Costs*; 2017.
32. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. 2009;6:1–6.
33. Peters BJM, Janssen V, Schramel FM, et al. Quantitative and qualitative assessment of real world data comparative effectiveness research of systemic therapies in lung oncology: a systematic review. *Cancer Epidemiol*. 2016;44:5–15.
34. Van Walraven C, Davis D, Forster AJ, et al. Time-dependent bias was common in survival analyses published in leading clinical journals. *J Clin Epidemiol*. 2004;57:672–682.

35. Johnson ML, Crown W, Martin BC, et al. Good research practices for comparative effectiveness research: analytic methods to improve causal inference from nonrandomized studies of treatment effects using secondary data sources: the ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report-Part III. *Value Health*. 2009;12:1062–1073.
36. Dreyer NA, Schneeweiss S, McNeil BJ, et al. GRACE principles: recognizing high-quality observational studies of comparative effectiveness. *Am J Manag Car*. 2010;16:467–471.
37. Suissa S. Immortal time bias in observational studies of drug effects. *Pharmacoepidemiol Drug Saf*. 2007;16:241–249.
38. Suissa S. Lower risk of death with SGLT2 inhibitors in observational studies: real or bias? *Diabetes Care*. 2018;41:6–10.
39. Hemkens LG, Contopoulos-Ioannidis DG, Ioannidis JP. Current use of routinely collected health data to complement randomized controlled trials: a meta-epidemiological survey. *CMAJ Open*. 2016;4:132–140.
40. Visvanathan K, Levit LA, Raghavan D, et al. Untapped potential of observational research to inform clinical decision making: American Society of Clinical Oncology research statement. *J Clin Oncol*. 2017;35:1845–1854.
41. Chien HC, Kao Yang YH, Bai JP. Trastuzumab-related cardiotoxic effects in Taiwanese women: a nationwide cohort study. *JAMA Oncol*. 2016;2:1317–1325.
42. De Glas NA, Hamaker ME, Kiderlen M, et al. Choosing relevant endpoints for older breast cancer patients in clinical trials: an overview of all current clinical trials on breast cancer treatment. *Breast Cancer Res Treat*. 2014;146:591–597.
43. Seruga B, Templeton AJ, Badillo FEV, et al. Under-reporting of harm in clinical trials. *Lancet Oncol*. 2016;17:209–219.
44. O'Sullivan AK, Thompson D, Drummond MF. Collection of health-economic data alongside clinical trials: is there a future for piggyback evaluations? *Value Health*. 2005;8:67–79.
45. Barcenas CH, Niu J, Zhang N, et al. Risk of hospitalization according to chemotherapy regimen in early-stage breast cancer. *J Clin Oncol*. 2014;32:2010–2017.
46. Schneider M, Zuckerman IH, Onukwugha E, et al. Chemotherapy treatment and survival in older women with estrogen receptor-negative metastatic breast cancer: a population-based analysis. *J Am Geriatr Soc*. 2011;59:637–646.
47. Yurkovich M, Avina-Zubieta JA, Thomas J, et al. A systematic review identifies valid comorbidity indices derived from administrative health data. *J Clin Epidemiol*. 2015;68:3–14.
48. Choi JY, Cho EY, Choi YJ, et al. Incidence and risk factors for congestive heart failure in patients with early breast cancer who received anthracycline and/or trastuzumab: a big data analysis of the Korean Health Insurance Review and Assessment service database. *Breast Cancer Res Treat*. 2018;171:181–188.
49. Nørgaard M, Ehrenstein V, Vandenbroucke JP. Confounding in observational studies based on large health care databases: problems and potential solutions–a primer for the clinician. *Clin Epidemiol*. 2017;9:185–193.
50. Iudici M, Porcher R, Riveros C, et al. Time-dependent biases in observational studies of comparative effectiveness research in rheumatology. A methodological review. *Ann Rheum Dis*. 2019;78(4):562–569.
51. Gavrielov-Yusim N, Friger M. Use of administrative medical databases in population-based research. *J Epidemiol Community Health*. 2014;68(3):283–287.
52. Langan SM, Schmidt SA, Wing K, et al. The reporting of studies conducted using observational routinely collected health data statement for pharmacoepidemiology (RECORD-PE). *BMJ*. 2018;363:1–19.
53. Carnahan RM. Mini-Sentinel's systematic reviews of validated methods for identifying health outcomes using administrative data: summary of findings and suggestions for future research. *Pharmacoepidemiol Drug Saf*. 2012;21:90–99.
54. Hassett MJ, Ritzwoller DP, Taback N, et al. Validating billing/encounter codes as indicators of lung, colorectal, breast, and prostate cancer recurrence using 2 large contemporary cohorts. *Med Care*. 2014;52(10):65–73.
55. Du XL, Osborne C, Goodwin JS. Population-based assessment of hospitalizations for toxicity from chemotherapy in older women with breast cancer. *J Clin Oncol*. 2002;20:4636–4642.
56. Schulman KL, Berenson K, Tina Shih YC, et al. A checklist for ascertaining study cohorts in oncology health services research using secondary data: report of the ISPOR oncology good outcomes research practices working group. *Value Health*. 2013;16:655–669.
57. Sanderson S, Tatt ID, Higgins JP. Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography. *Int J Epidemiol*. 2007;36:666–676.
58. Berger ML, Sox H, Willke RJ, et al. Good practices for real-world data studies of treatment and/or comparative effectiveness: recommendations from the joint ISPOR-ISPE Special Task Force on real-world evidence in health care decision making. *Pharmacoepidemiol Drug Saf*. 2017;26:1033–1039.
59. Dreyer NA, Velentgas P, Westrich K, et al. The GRACE checklist for rating the quality of observational studies of comparative effectiveness: a tale of hope and caution. *J Manag Care Spec Pharm*. 2014;20:301–308.