Journal of Clinical Epidemiology 109 (2019) 125–132

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

Real-world evidence was feasible for estimating effectiveness of chemotherapy in breast cancer: a cohort study

Ewan Gray[a,*], Joachim Marti[b], David H. Brewster[a], Jeremy C. Wyatt[c], Romain Piaget-Rossel[b], Peter S. Hall[a]

[a]*Edinburgh Clinical Trials Unit, University of Edinburgh, Edinburgh, UK*
[b]*Institut universitaire de médecine sociale et préventive, University of Lausanne, Lausanne, Vaud, Switzerland*
[c]*Wessex Institute, University of Southampton, Southampton, UK*

Accepted 16 January 2019; Published online 31 January 2019

Abstract

Objective: Evidence-based guidelines recommend adjuvant chemotherapy in early stage breast cancer whenever treatment benefit is considered sufficient to outweigh the associated risks. However, many groups of patients were either excluded from or underrepresented in the clinical trials that form the evidence base for this recommendation. This study aims to determine whether using administrative health care data—real world data—and econometric methods for causal analysis to provide "real world evidence" (RWE) are feasible methods for addressing this gap.

Methods: Cases of primary breast cancer in women from 2001 to 2015 were extracted from the Scottish cancer registry (SMR06) and linked to other routine health records (inpatient and outpatient visits). Four methods were used to estimate the effect of adjuvant chemotherapy on disease-specific and overall mortality: (1) regression with adjustment for covariates, (2) propensity score matching, (3) instrumental variables analysis, and (4) regression discontinuity design. Hazard ratios for breast cancer mortality and all-cause mortality were compared to those from a meta-analysis of randomized trials.

Results: A total of 39,805 cases were included in the analyses. Regression adjustment, propensity score matching, and instrumental variables were feasible, whereas regression discontinuity was not. Effectiveness estimates were similar between RWE and randomized trials for breast cancer mortality but not for all-cause mortality.

Conclusions: RWE methods are a feasible means to generate estimates of effectiveness of adjuvant chemotherapy in early stage breast cancer. However, such estimates must be interpreted in the context of the available randomized evidence and the potential biases of the observational methods.

Keywords: Real world evidence; Propensity score; Instrumental variables; Regression discontinuity; Feasibility; Meta-analysis; Breast cancer; Adjuvant chemotherapy

1. Introduction

Adjuvant chemotherapy is indicated in cases of early stage breast cancer whenever treatment benefit is considered sufficient to outweigh the associated risks [1]. The benefits of chemotherapy, in terms of improved survival and disease-free survival, are known to vary on a case by case basis depending on a number of prognostic markers. Likewise, the risks of chemotherapy vary depending on the characteristics of the patient [2]. Patients, with advice from their clinicians, must choose whether or not to undergo adjuvant chemotherapy taking account of these individual factors and their own beliefs and preferences. The shared decision-making process requires a patient-specific risk assessment and the effective communication of patient preference, risks, and benefits information between the clinician and the patient. To facilitate this process and help improve decisions, a number of tools have been developed that quantify the risks and benefits of the available treatments [3–5]. Data from randomized controlled trials (RCTs) are the primary source for decision-making tools; however, these are limited by strict patient inclusion criteria, leading to concerns about whether treatment benefit estimates are accurate for all patients (generalizability). It is proposed that real world evidence (RWE) can contribute to informing these decisions by providing accurate

Funding: This study was funded by the Chief Scientist Office (CSO), Scotland. The funder had no role in the design or reporting of the study.

Conflict of interest statement: All authors have no conflicts of interest to declare.

* Corresponding author. Edinburgh Clinical Trials Unit, Nine Edinburgh Bioquarter, 9 Little France Road, Edinburgh EH16 4UX, UK. Tel.: +44 01316519944; fax: +44 01315373851.

E-mail address: ewan.gray@ed.ac.uk (E. Gray).

https://doi.org/10.1016/j.jclinepi.2019.01.006
0895-4356/

What is new?

Key findings

- Regression adjustment, propensity score matching, and instrumental variables were feasible methods for estimating the effectiveness of adjuvant chemotherapy in early stage breast cancer, whereas regression discontinuity design was not.

What this adds to what was known?

- Estimates of treatment effectiveness were similar between real world evidence (RWE) methods and a meta-analysis of randomized trials for breast cancer mortality but not for all-cause mortality.

What is the implication and what should change now?

- RWE should be interpreted cautiously, in the context of the available RCT evidence and with consideration of alternative methods that can be implemented using observational data.

treatment benefit estimates from more representative real world data [6].

Real world data (RWD) refer to data used for decision making that are collected outside of randomized controlled trials (RCTs) [7]. RCTs are the gold standard for reliably measuring treatment efficacy and constitute the primary source of evidence to inform decision making in health care. Randomization, correctly implemented [8], guarantees unbiased estimates of the average treatment effect (in expectation) by ensuring balance of observed and unobserved covariates in treatment and control groups. Real world data in contrast are observational in nature and therefore subject to additional sources of bias [9]. Methods of analysis, ''designs,'' have been developed to allow less biased estimates of treatment effects under reasonable assumptions. These methods have contributed to the ''credibility revolution'' in economics [10] and have led some to re-evaluate existing hierarchies of evidence in medicine [11].

The ''quasi-experimental'' methods available to researchers differ in the mechanisms used to mirror random assignment; historically the greatest limitations to their application have been data availability and quality and concerns about the feasibility of more advanced methods.

This study makes use of high quality routine data to implement alternative candidate methods and compare estimates both between methods and with the available RCT evidence. The randomized evidence comes from a series of progressively updated meta-analyses published by the Early Breast Cancer Trialists Collaborating Group (EBCTCG) [2,12,13]. Results from the most recent meta-analysis are used to make comparisons with the RWE estimates from this study.

1.1. We consider four candidate methods

1 Regression with adjustment for covariates (RA): Uses multiple regression-based methods to adjust for the imbalance in observed covariates between treated and untreated cases.
2 Propensity score matching (PSM) [14]: Uses rich prognostic data to create propensity scores and match treated and untreated cases, reducing confounding by indication.
3 Instrumental variables (IV) [15]: Makes use of variables that are assumed to causally affect the treatment decision but have no effect on outcomes other than indirectly, via changing the probability of treatment.
4 Regression discontinuity design (RDD) [16]: Exploits the variation in treatment use created by a treatment guideline based on a threshold level of estimated treatment benefit provided by an online tool.

This study aims to explore the feasibility and compare the results of RWE methods for estimating the effectiveness of adjuvant chemotherapy for early stage breast cancer. Because of the reasonable concerns about bias in RWE methods, the approach we have taken is exploratory in nature and seeks to provide extensive contextual information to inform the judgments of readers on the interpretation of RWE estimates. We believe such an open approach, which differs from the typical "stand alone" inference of a randomized trial, is necessary if RWE is to be useful for informing patient and clinician decision making in this setting.

A key feature of this study is that all methods make use of prognostic and treatment benefit predictions about individual women provided by an online prognostication and treatment benefit tool for patients with early stage breast cancer—PREDICT [5].

Some RWE methods have previously been employed in this setting. PSM was used with a large observational study conducted in the USA [17], comparing mastectomy and breast conserving surgery in node-negative patients using a registry data set. The results corresponded closely with previously reported trials and provided evidence that the estimated hazard ratios could be generalized beyond trial populations, and this information was influential for clinical practice. The success of PSM for comparing surgical strategies in the same patient group is one reason to believe PSM might also be appropriate for addressing questions relating to the effectiveness of adjuvant therapies. Other PSM studies focusing on adjuvant therapies have not made comparisons with randomized data.[1]

Application of RDD for the evaluation of health care interventions has recently received increased attention, and

[1] PSM has also previously been applied to estimate adjuvant chemotherapy effectiveness in a large case series from a single institution in France [18]. PSM has also been used for estimating the effects of adjuvant chemotherapy for older women [19,20] (a specific trial ineligible group) in data from the USA.

Table 1. Use of person-specific PREDICT scores in the four RWE methods

Method	How PREDICT outputs were used:
Regression adjustment	PREDICT Prognostic score and benefit score were used as explanatory variables. This effectively includes prognostic variable interactions and transforms informed by external evidence (from derivation cohort and previous studies that informed the specification used in PREDICT modeling).
Propensity score matching	Prognostic score and benefit scores were used as explanatory variables in the propensity score model.
Instrumental variables	PREDICT benefit score was used as an instrument. Benefit score interacted with post-2010 dummy variable used as an alternative instrument. Prognostic score is an (independent) explanatory variable.
Regression discontinuity	Benefit score used as the continuous assignment variable

there are some successful examples of this method [18]. However, the application of RDD to this clinical area is, to the best of our knowledge, completely novel.

2. Methods

2.1. Patient data

Patient level data were transferred into the National Services Scotland Safe Haven as an extract from the Scottish Cancer Registry (SCR). All records in the registry with a diagnosis of primary invasive breast cancer (ICD-10 C50), diagnosed in the period between January 2001 and December 2015, were retrieved for analysis. SCR is a population-based registry that covers all residents of Scotland (population approximately 5.5 million). National Records of Scotland provides notification of deaths for registry records. Vital status was recorded up to 1st February 2017 in the analysis extract.[2] Deaths due to breast cancer were defined in accordance with the ICD-10 coding system for causes of death, recorded either as the primary cause of death or as one of three contributing causes of death. Data were restricted to the first occurrence of a primary breast cancer for each patient; subsequent, primary breast cancers were excluded. Data linkage was provided by Information Services Division (ISD) to Scottish hospital inpatient and day case records (SMR01) and outpatient records (SMR00). Deterministic linkage was achieved using the Community Health Index number unique individual identifiers, which includes a check digit. The linked data sets included all records linked to an included registry case from the period up to 5 years before the date of diagnosis. Prognostic factors available for use in the analysis, including the derived PREDICT scores, are described in detail in the supplementary appendix (SA1).

Details of the use of PREDICT scores in each method are displayed in Table 1. In the RA, PSM and IV methods PREDICT version 2 scores were used. "Prognostic score" was the PREDICT probability of death from any cause over 10 years. In models with breast cancer-specific mortality as the outcome, this was the probability of death because of breast cancer over 10 years. PREDICT benefit score was the difference in the probability of survival at 10 years following diagnosis with and without adjuvant chemotherapy.

[2] For those emigrating from Scotland a date of embarkation is available allowing censoring these observations at the appropriate time.

Cases were excluded if the patient was male, had advanced cancer (clinical M stage = 1), did not receive surgery, or received neoadjuvant therapy (chemotherapy or hormone therapy recorded before surgery).

Because some groups of patients, such as over 70s and those with comorbidities, were either excluded from or underrepresented in the randomized trials, we made a comparison with a subgroup of RWE patients who would meet trial inclusion criteria as well as with the full cohort. Cases were defined as "trial represented" (TR) if they were under 70 years of age with no recorded Charlson comorbidities and met criteria related to prognostic factors; either node positive or with 2 or more of: [1] >30 mm tumor size, [2] grade 3, [3] ER-, and [4] Her2+ status. The definition of trial represented was based on assessment of the protocols of a number of trials in the meta-analysis [2] and clinical expert opinion. It should be noted that trial inclusion/exclusion protocols varied to some degree and often included elements of clinical judgment.

2.2. Econometric methods

All analyses were repeated for two outcomes: breast cancer-specific mortality and all-cause mortality as recorded on death certificates. Furthermore, the analyses were repeated in the full cohort and the TR subgroup. Each method of analysis therefore produced four estimated hazard ratios. The details of the implementation are reported in the supplementary appendix (SA2).

2.3. Comparison with randomized studies

Clinical expert opinion suggests that the predominant chemotherapy regimens in use in Scotland during the 2001–2015 period were newer anthracycline-containing regimens (CAF/CEF); therefore, estimates for these types of regimen were used as the best comparison with the RWE estimates. Both direct evidence, from trials comparing newer anthracycline-containing regimens vs. placebo, and indirect evidence via trials of newer anthracycline-containing regimens vs. other active regimens were considered. This required some reanalysis using the trial level summary statistics presented in [2] as a network meta-analysis [19]. In the primary analysis vs. RWE methods, the comparison was made using only direct trial evidence.

To compare RWE with randomized evidence, two approaches were taken. First, a statistical test of the difference

in the estimated treatment effect between trial and observational sources based on z scores, as suggested by Ioannidis et al. [20]. z scores above 1.96 or below −1.96 are considered as sufficient evidence that the difference in estimates is beyond that expected by chance. Second, meta-analysis estimates were calculated with and without inclusion of RWE estimates. Results were presented in forest plots to allow visual assessment of heterogeneity. A random effects model was assumed because of known between-trial and trial-observational differences in study populations. Cochrane's Q was used as a statistical test of heterogeneity.

3. Results

3.1. Sample selection

A total of 63,116 records were retrieved from the registry. Following removal of duplicate records (non-first cancers and multiple synchronous tumors) and application of the exclusion criteria, a total of 39,805 cases remained in the primary analysis. The process is detailed in Figure 1. Thirteen percentage of otherwise eligible cases contained missing prognostic variable data, but in most cases only a single variable was missing. A summary of sample characteristics can be found in Table 2.

3.2. Feasibility

RA using Cox regression was feasible in both TR and full cohort groups. PSM was also determined to be feasible. The distribution of propensity scores is displayed in supplementary appendix (SA3). The region of common support was sufficient in both the trial represented and full cohort groups. The balance of baseline covariates between matched treated and nontreated units in both groups was compared for each matching method (SA3). PSM 1 showed some imbalances for the TR group and more severe imbalance in the full cohort. PSM 2 achieved good balance in the TR but not in the full cohort. PSM 3 achieved good balance for both TR and full cohorts. Consequently, PSM 3 was selected as the primary analysis. Results for all PSM methods are available in appendix SA3.

Both IV approaches demonstrated feasibility. The first-stage regression results are displayed in appendix SA4. Both instruments showed promise through statistically significant associations with chemotherapy use in the first stage regressions. Note that much of the variation in chemotherapy use caused by the instruments may be in node-negative patients who would not meet the defined trial represented criteria; therefore, these instruments may be more powerful in the full cohort.

RDD was not feasible. Inspection of histograms confirmed that the requirement of continuity in the region of the 3% and 5% thresholds was met for PREDICT v2 but not v1.2 (Figs. SA5.1 and SA5.2), eliminating PREDICT 1.2 as a candidate assignment variable.

The PREDICT version 2 benefit score with 3% and 5% thresholds had potential for an RDD. To determine whether or not treatment guidelines and norms actually create such a discontinuity, we inspected the binned scatterplots

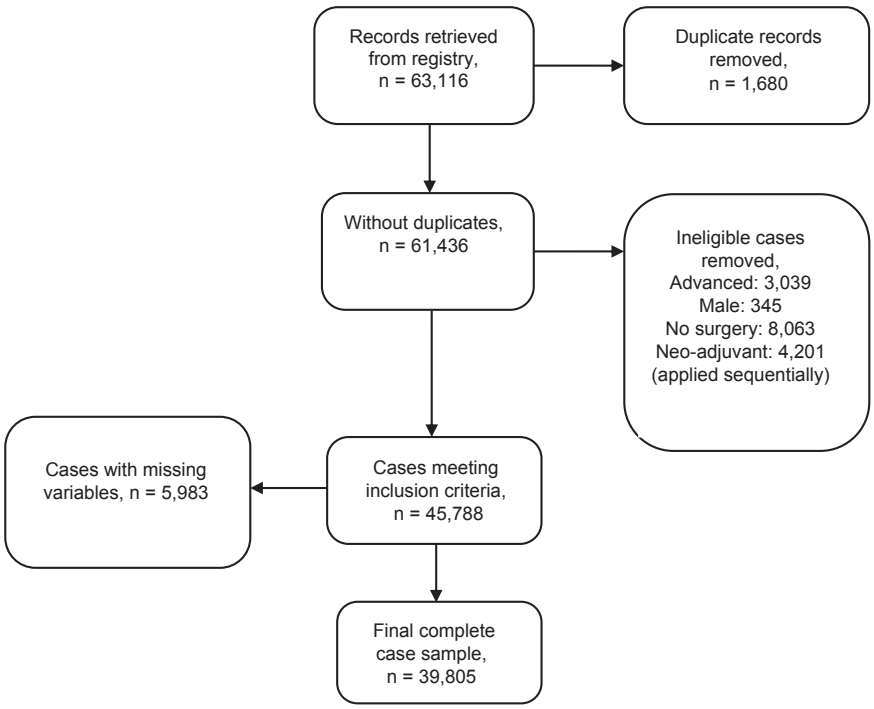

Fig. 1. Sample selection flow diagram.

Table 2. Sample characteristics

Statistic	Full cohort	Trial representative cohort
Total number of subjects	39,805	12,870
Total time at risk (yr)	278,984	93,222
Median follow-up (yr)	6.35	6.64
Number of breast cancer deaths	4,977	2,126
Number of other deaths	3,624	511
Five-year breast cancer survival rate	87.1%	86.8%
Median age at diagnosis, years	61	54

Variable	Number	(%)	Number	(%)
Age <35	503	1.3	321	2.5
35–49	7,015	17.6	3,879	30.1
50–64	16,792	42.2	6,655	51.7
65–74	9,797	24.6	2,015	15.7
≥75	5,698	14.3	0	0
Nodes 0	26,275	66	3,294	25.6
Nodes 1	5,747	14.4	4,154	32.3
Nodes 2–4	4,456	11.2	3,202	24.9
Nodes 5–9	1,635	4.1	1,079	8.4
Nodes 10+	1,453	3.7	978	7.6
Tumor size <10 mm	5,453	13.7	717	5.6
10–19 mm	15,774	39.6	3,907	30.4
20–29 mm	10,727	26.9	4,029	31.3
30–49 mm	5,967	15	3,146	24.4
≥50 mm	1,884	4.7	1,071	8.3
Grade I	5,866	14.7	758	5.9
II	19,130	48.1	4,473	34.8
III	14,606	36.7	7,564	58.8
ER-	6,208	15.6	3,900	30.3
ER+	33,597	84.4	8,970	69.7
Chemotherapy	14,589	36.7	10,439	81.1
Hormone therapy	29,991	75.3	7,822	60.8
Chemo + hormone therapy	8,875	22.3	6,060	47.1
Screen detected	14,887	37.4	3,602	28
Symptomatic	24,827	62.4	9,205	71.5
Charlson Index ≥1	2,455	6.2	0	0
Charlson Index = 0	37,350	93.8	12,870	100
SIMD deprivation quintile 1 (most deprived	6,893	17.3	2,377	18.5
SIMD 2	7,823	19.7	2,480	19.3
SIMD 3	8,478	21.3	2,705	21
SIMD 4	8,300	20.9	2,719	21.1
SIMD 5 (least deprived)	8,310	20.9	2,588	20.1
<3% Chemo benefit PREDICT	24,475	61.5	3,357	26.1
3–5%	8,382	21.1	4,570	35.5
>5%	6,948	17.5	4,943	38.4

Variable	Mean	s.d.	Mean	s.d.
Age	60.72	0.067	53.47	0.089
Tumor size	22.03	0.076	27.02	0.149

(*Continued*)

Table 2. Continued

Variable	Mean	s.d.	Mean	s.d.
Total inpatient days	2.88	0.065	1.35	0.061
Total outpatient visits	6.2	0.048	5.33	0.07
PREDICT benefit score	2.93	0.012	4.66	0.02

s.d., standard deviation.

displayed in Figures SA5.3 and S5.4, to visualize the relationship between the assignment variable and the probability of chemotherapy use.

The binned scatterplots show no clear discontinuity in the probability of using chemotherapy at the 3% or 5% thresholds in the trial-represented group. The plots suggest that the probability of chemotherapy use is already high by 3% chemotherapy benefit and increases only gradually past this threshold. Based on these core assumptions not being met, estimation of the treatment effect using RDD was halted.

3.3. Comparison with trial meta-analysis estimates

Full details of the randomized trial evidence and network meta-analysis results are in the Supplementary Appendix (SA6). Comparisons are made with only the direct randomized evidence to simplify the analysis and presentation and because the network meta-analysis results indicate little difference from including the indirect evidence. All the preceding RWE estimates of the effectiveness of adjuvant chemotherapy are summarized together in Table 3, together with the combined estimates from a random effects meta-analysis and results of a test of the difference in estimated HR between RWE and trial meta-analysis. Forest plots displaying the estimated HRs for individual trials and RWE methods are displayed in Figure 2. Forest plots displaying the individual and pooled estimates are available in SA6.

Breast cancer mortality estimates from each RWE method lie relatively close to the pooled randomized trial estimates, except for IV1. In contrast, there is some evidence of differences between RWE and trial evidence in relation to all-cause mortality, with the PSM estimating lower hazard ratios than trials.

For all RWE methods except PSM Cox, estimates of the hazard ratios for all-cause mortality are lower than for BC-specific mortality, in contrast to the trial estimates. This suggests a downward bias in the all-cause mortality estimates, as the effect of chemotherapy on all-cause mortality is expected to be smaller and proportional to its effect on BC mortality. This is based on the mechanism of action and was demonstrated in the randomized trial meta-analysis results.

Some differences between RWE estimates are large enough to produce differences in the combined HRs which may be of clinical importance. Estimates from the IV analyses are relatively imprecise in the TR group, which may

Table 3. Comparison of effect sizes (HR, 95% confidence intervals) and pooled estimates from randomized trials, RWE (trial represented and RWE full cohort)

	Breast cancer mortality		All-cause mortality	
Source of estimates	**HR**	**Combined HR**	**HR**	**Combined HR**
EBCTCG meta-analysis: direct evidence newer anthracycline regimens vs. placebo				
	0.71 (0.62; 0.83)	–	0.83 (0.73; 0.94)	–
RWE estimates (trial represented)				
RA	0.76 (0.67; 0.87)	0.74 (0.67; 0.82)	0.69[b] (0.62; 0.77)	0.75 (0.69; 0.81)
PSM Cox	0.69 (0.57; 0.84)	0.71 (0.63; 0.79)	0.71 (0.61; 0.83)	0.78 (0.71; 0.86)
PSM LR	0.77 (0.64; 0.93)	0.74 (0.66; 0.82)	0.73 (0.63; 0.85)	0.79 (0.71; 0.87)
IV1	0.91 (0.70; 1.20)	0.75 (0.66; 0.86)	0.78 (0.61; 1.00)	0.82 (0.73; 0.91)
IV2	0.85 (0.64; 1.13)	0.74 (0.65; 0.84)	0.74 (0.57; 0.95)	0.81 (0.72; 0.91)
RWE estimates (full cohort)				
RA	0.82 (0.75; 0.89)	0.79 (0.73; 0.85)	0.77 (0.72; 0.83)	0.78 (0.74; 0.83)
PSM Cox	0.66 (0.59; 0.74)	0.68 (0.62; 0.75)	0.67[b] (0.61; 0.74)	0.75 (0.67; 0.85)
PSM LR	0.68 (0.61; 0.76)	0.69 (0.63; 0.76)	0.67[b] (0.61; 0.73)	0.76 (0.67; 0.85)
IV1	0.88[a] (0.75; 1.02)	0.78 (0.68; 0.89)	0.82 (0.73; 0.92)	0.82 (0.75; 0.90)
IV2	0.82 (0.71; 0.96)	0.76 (0.69; 0.85)	0.81 (0.72; 0.91)	0.82 (0.75; 0.89)

[a] *P*-value < 0.1.

[b] *P*-value < 0.05. *P*-values obtained from a student t-test with H0: RWE hazard ratio = EBCTCG hazard ratio. Forest plot of EBCTCG estimates and each RWE estimate obtained from trial represented sample are provided in SA6 Figures SA6.4-SA6.23.

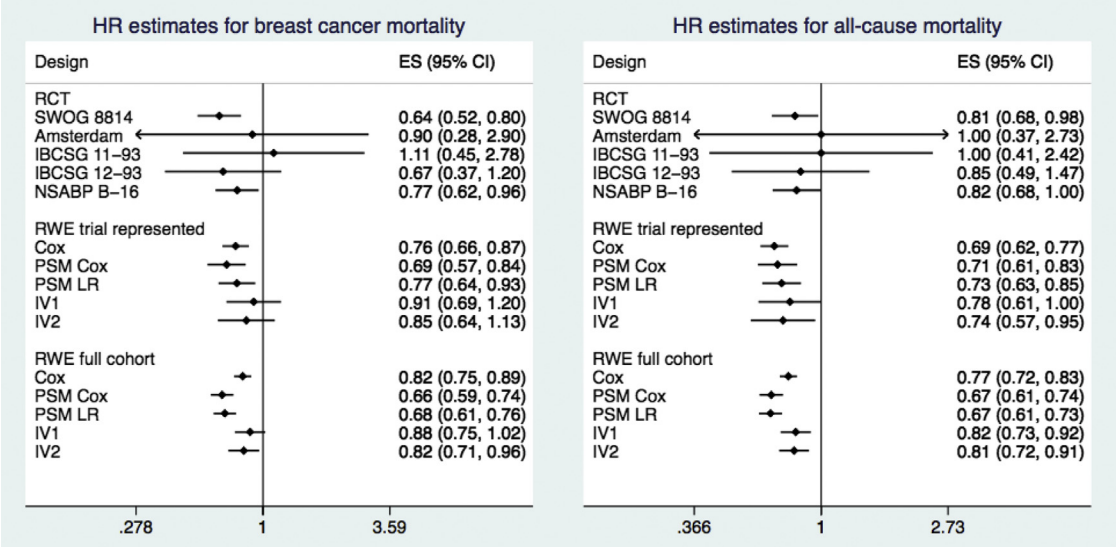

Fig. 2. Forest plots; Trial and RWE estimated HR for breast cancer and all-cause mortality.

be due to the relatively small proportion of variation in the use of chemotherapy explained by the instruments. Despite our relatively large sample sizes, the synthesis of RWE and trial evidence in a random effects meta-analysis reduces the uncertainty about the HR by only a small amount compared to using the randomized trial data alone, irrespective of using the TR or full cohort samples.

4. Discussion

The RWD used in this study is of large size and high quality. Linkage to a range of other routine data sets allows an assessment of patient comorbidities that would not be possible with registry records alone. Use of an existing, validated prognostic score efficiently uses prognostic information in a manner consistent with existing epidemiological evidence.

We have attempted to be comprehensive and transparent in the application and evaluation of RWE methods, rather than simply presenting one method with what we judge the ‘‘best’’ result or justification. This guards against the potential for spurious results arising from a selective post hoc application of the methods or selective use of a particular specification within each method based on the results. Using RWD, there is clearly a wider scope for creating bias in this way, as compared to using trial data. This is an important analytical problem similar to ‘‘p-hacking’’ or a ‘‘garden of forking paths’’ that has been described in other settings [21].

A limitation of this study is that the comparison of real world and clinical trial evidence may not directly assess the validity of the real-world evidence. Validating against clinical trials could be misleading if the average treatment effect in a trial represented RWD group and the actual trial samples differs systematically. At the same time, interpreting the results as estimates of systematic differences in treatment effects between trial and real-world populations will be unwise if the observational methods lack internal validity. In this study, the use of a trial-represented population in the RWD should minimize this potential problem. Other solutions to this problem could be attempted in future research if additional data are available. Although we were fortunate that a comprehensive IPD meta-analysis was available, this analysis was restricted to using the summary statistics from the published report. This limited the degree to which we could match the randomized trial and RWD populations for baseline characteristics and treatment protocols or make use of other methods of adjustment [22]. Future studies should consider a deeper collaboration with trialists to allow more nuanced comparison. Another area that could be explored in future research is the design of other tests of internal validity, such as falsification endpoints [23], of RWE methods such as IV in this setting.

This analysis would be enhanced if more detailed data were available for specific variables. Chemotherapy use was only available as a binary variable, rather than the details of the regimens used. Her2 status and trastuzumab use were not available for many observations (and not available in any case before 2009). Thirteen percentage of cases were excluded from the analysis because of other missing prognostic data. Another limitation of this analysis is that it relies on accurate coding of causes of death on death certificates to ascertain breast cancer deaths. This may be less accurate in RWD than in the trial setting, where detailed follow-up to ascertain deaths with recurrence of breast cancer has been used. RWE could also be enhanced by reporting of effects on multiple alternative causes of death (e.g., breast cancer, cardiovascular, other) where such data are available, using an appropriate methodology to account for competing risks. This would also more nuanced conclusions regarding the potential biases related to each of these outcomes.

5. Conclusions

Regression adjustment, PSM, and IV were feasible methods for obtaining treatment effect estimates from these real-world data, whereas RDD was not. Although concordance with randomized trial evidence was demonstrated for cause-specific mortality, there was some indication of bias in estimates of chemotherapy effects on all-cause mortality. We conclude that even with large, good quality data sets and careful validation of the assumptions underlying these methods, RWE should be interpreted cautiously, in the context of the available RCT evidence and with consideration of alternative methods that can be implemented using observational data.

CRediT authorship contribution statement

Ewan Gray: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing - original draft, Project administration. **Joachim Marti:** Conceptualization, Methodology, Writing - review & editing. **David H. Brewster:** Conceptualization, Methodology, Writing - review & editing. **Jeremy C. Wyatt:** Conceptualization, Methodology, Writing - review & editing. **Romain Piaget-Rossel:** Methodology, Software, Formal analysis, Writing - review & editing. **Peter S. Hall:** Conceptualization, Methodology, Writing - review & editing, Supervision.

Acknowledgments

The authors thank the SATURNE advisory group—David Cameron, Fiona Watt, Iain MacPherson, Larry Hayward, Colin McCowen, Gianluca Baio, and Paul Pharoah—for their generous advice and support. The authors also thank David MacAllister, Rachel Meacock, Patrick Taffe, the HESG participants, and two anonymous reviewers for their helpful suggestions. This study was funded by the Chief Scientist Office (CSO, HIPS/16/26). The funder had no role in the design or reporting of the study. Finally, the authors thank all the women who have provided the data that has been used in this study.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclinepi.2019.01.006.

References

[1] National Collaborating Centre for Cancer. CG80 Early and locally advanced breast cancer: diagnosis and treatment. London, UK: National Institute For Health and Clinical Excellence; 2009.
[2] Early Breast Cancer Trialists' Collaborative Group. Comparisons between different polychemotherapy regimens for early breast cancer: meta-analyses of long-term outcome among 100 000 women in 123 randomised trials. Lancet 2012;379(9814):432−44.
[3] Galea MH, Blamey RW, Elston CE, Ellis IO. The nottingham prognostic Index in primary breast-cancer. Breast Cancer Res Treat 1992;22(3):207−19.
[4] Ravdin PM, Siminoff LA, Davis GJ, Mercer MB, Hewlett J, Gerson N, et al. Computer program to assist in making decisions about adjuvant therapy for women with early breast cancer. J Clin Oncol 2001;19:980−91.
[5] Wishart GC, Azzato EM, Greenberg DC, Rashbass J, Kearins O, Lawrence G, et al. PREDICT: a new UK prognostic model that predicts survival following surgery for invasive breast cancer. Breast Cancer Res 2010;12(1):R1.
[6] Sherman RE, Anderson SA, Dal Pan GJ, Gray GW, Gross T, Hunter NL, et al. Real-world evidence - what is it and what can it tell us? N Engl J Med 2016;375(23):2293−7.
[7] Garrison LP Jr, Neumann PJ, Erickson P, Marshall D, Mullins CD. Using real-world data for coverage and payment decisions: the ISPOR real-world data task force report. Value Health 2007;10(5):326−35.
[8] Altman DG, Schulz KF, Moher D. Turning a blind eye - testing the success of blinding and the CONSORT statement. Br Med J 2004; 328(7448):1135.
[9] Rosenbaum PR. Discussing hidden bias in observational studies. Ann Intern Med 1991;115:901−5.
[10] Angrist JD, Pischke J-S. The credibility revolution in empirical economics: how better research design is taking the con out of econometrics. J Econ Perspect 2010;24(2):3−30.
[11] Craig P, Cooper C, Gunnell D, Haw S, Lawson K, Macintyre S, et al. Using natural experiments to evaluate population health interventions: new medical research council guidance. J Epidemiol Community Health 2012;66:1182−6.
[12] Early Breast Cancer Trialists' Collaborative Group. Polychemotherapy for early breast cancer: an overview of the randomised trials. Lancet 1998;352(9132):930−42.
[13] Early Breast Cancer Trialists' Collaborative Group. Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. Lancet 2005;365(9472):1687−717.
[14] Rosenbaum P, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika 1983;70(1): 41−55.
[15] Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. J Am Stat Assoc 1996;91:444−55.
[16] Oldenburg CE, Moscoe E, Barnighausen T. Regression discontinuity for causal effect estimation in epidemiology. Curr Epidemiol Rep 2016;3:233−41.
[17] General Accounting Office (GAO) United States of America. Breast conservation versus mastectomy. In: Patient survival in day-to-day medical practice and in randomized studies. Washington, DC: General Accounting Office; 1994.
[18] Moscoe E, Bor J, Barnighausen T. Regression discontinuity designs are underutilized in medicine, epidemiology, and public health: a review of current and best practice. J Clin Epidemiol 2015;68:122−33.
[19] White IR, Barrett JK, Jackson D, Higgins JPT. Consistency and inconsistency in network meta-analysis: model estimation using multivariate meta-regression. Res Synth Methods 2012;3(2):111−25.
[20] Ioannidis JP, Haidich AB, Pappa M, Pantazis N, Kokori SI, Tektonidou MG, et al. Comparison of evidence of treatment effects in randomized and nonrandomized studies. JAMA 2001;286:821−30.
[21] Simmons JP, Nelson LD, Simonsohn U. False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. Psychol Sci 2011;22(11):1359−66.
[22] Signorovitch J, Sikirica V, Erder M, Xie J, Lu M, Hodgkins P, et al. Matching-adjusted indirect comparisons: a new tool for timely comparative effectiveness research. Value Health 2012;15(6):940−7.
[23] Prasad V, Jena AB. Prespecified falsification end points: can they validate true observational associations? JAMA 2013;309:241−2.