BMJ Open

Can analyses of electronic patient records be independently and externally validated? Study 2: the Effect of Beta-Adrenoceptor Blocker Therapy on Cancer Survival; a Retrospective Cohort Study

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2014-007299
Article Type:	Research
Date Submitted by the Author:	01-Dec-2014
Complete List of Authors:	Springate, David; University of Manchester, Centre for Primary Care Ashcroft, Darren; University of Manchester, Centre for Pharmacoepidemiology Kontopantelis, Evangelos; University of Manchester, Institute of Population Health Doran, Tim; University of York, Health Sciences Ryan, Ronan; University of Birmingham, Reeves, David; University of Manchester, Institute of Population Health; Centre for Biostatistics
Primary Subject Heading:	General practice / Family practice
Secondary Subject Heading:	Research methods, Evidence based practice, Health informatics, Oncology, Pharmacology and therapeutics
Keywords:	PRIMARY CARE, Health informatics < BIOTECHNOLOGY & BIOINFORMATICS, ONCOLOGY, STATISTICS & RESEARCH METHODS

Can analyses of electronic patient records be independently and externally validated? Study 2: the Effect of Beta-Adrenoceptor Blocker Therapy on Cancer Survival; a Retrospective Cohort Study

Corresponding author
David Reeves
NIHR School for Primary Care Research, Centre for Primary Care, Institute of Population Health, University of Manchester, Manchester M13 9PL, UK
Email: david.reeves@manchester.ac.uk
Tel: 0161 275 3536

Authors in order
David Springate, *research associate*1 2, Darren M Ashcroft, *professor*4, Evangelos Kontopantelis, *senior research fellow*1 2, Tim Doran, *professor5*, Ronan Ryan *research fellow6*, David Reeves, *reader*1 3

1 NIHR School for Primary Care Research, Centre for Primary Care, Institute of Population Health, University of Manchester
2 Centre for Health Informatics, Institute of Population Health, University of Manchester
3 Centre for Biostatistics, Institute of Population Health, University of Manchester
4 Centre for Pharmacoepidemiology and Drug Safety Research, Manchester Pharmacy School, University of Manchester
5 Department of Health Sciences, University of York
6 Primary Care Clinical Sciences, School of Health and Population Sciences, University of Birmingham

Keywords
Validity of electronic health records; primary care; health informatics; drug effectiveness; cohort study

Word count
4563

Abstract

Objectives To conduct a fully independent and external validation of a research study based on one electronic health record database, using a different database sampling the same population.

Design Retrospective cohort analysis of the effects of beta-blocker therapy on all-cause mortality in cancer patients.

Setting Electronic health record databases (Clinical Practice Research Datalink (CPRD) and Doctors' Independent Network (DIN)) of longitudinal patient consultation data from general practices distributed throughout the UK.

Participants CPRD data for 11302 patients with solid cancers compared to previously published results from DIN for 3462 patients. Study period January 1997 to December 2006.

Primary and secondary outcome measures All-cause mortality: overall; by treatment subgroup and by cancer-site.

Results Using CPRD, beta-blocker use was not associated with cancer mortality (hazard ratio=1.03, 95%CI 0.93 to 1.14), but in DIN beta-blocker users had significantly higher mortality (hazard ratio = 1.18, 95%CI 1.04 to 1.33). However, these rates were not statistically different (p=0.063), but did differ for patients on beta-blockers alone (p<0.001). Study-specific results for nine individual cancer sites were quite different, but under direct comparison differed significantly for prostate and pancreas cancers only. Results were robust under sensitivity analyses for differences in sample characteristics, but we could not be certain that mortality was identically defined in both databases.

Conclusions We found a complex pattern of similarities and differences between databases. Overall treatment effect estimates were not statistically different and the main clinical conclusions did not differ, though some subgroup effects differed significantly. The present study together with previous independent and non-independent replication studies, constitutes a growing body of evidence that when analysed to a common protocol, a variety of different UK PCDs produce effect estimates that are comparable within statistical limits of accuracy, though confirmatory studies are still advised, especially regarding small treatment effects and subgroup results.

Strengths and limitations of this study

- Drug effectiveness studies applying the same analysis protocol to different electronic health record (EHR) databases have typically compared EHRs covering different patient populations or replications have not been independently conducted. This paper reports on a fully independent validation of a published EHR-based study using a different EHR database sampling from the same underlying population.

- Despite purporting to cover the same general UK population, there were some notable demographic and clinical differences between the CPRD and DIN cancer cohorts. Sensitivity analysis indicated that these had only a minimal effect on treatment effect estimates, but we were unable to account for a difference in mortality rates between the cohorts.

- Examined separately, the CPRD- and DIN-based studies produced quite different pictures of the risks of beta-blockers overall and in relation to different cancer types. But when directly compared, except for a few specific subgroup results, estimates of treatment effect did not differ statistically and the principal clinical conclusion was the same.

- The present study adds to evidence from our previous independent replication study and other non-independent replications, that the application of identical analytical methods to a variety of different UK primary care databases produces treatment effect estimates that are in most respects comparable.

INTRODUCTION

Large-scale electronic health record databases (EHRs) are widely regarded as an important new tool for medical research. The major UK "primary care databases" (PCDs) are some of the largest and most detailed sources of electronic patient data available, holding detailed long-term clinical data for many millions of patients. Researchers are increasingly using these resources [1], which provide a means for researching questions in primary care that cannot feasibly be addressed by other means, including unintended consequences of drug interventions, where ethical considerations, the required numbers of patients, or length of follow-up can make an RCT impractical.

Concerns remain, however, about the validity of studies based on such data, including uncertainties about data quality, data completeness, and the potential for bias due to both measured and unobserved confounders. Most work on EHR validity has focused on the accuracy or completeness of the individually recorded data values, such as consultation recording [2], disease diagnoses [3,4] and risk factors [5-7]. Another approach to testing the validity of EHR-based studies is to compare the results to those obtained from equivalent investigations conducted on other, independent, datasets. Agreement of results helps to reassure that the findings do not depend upon the source of the data, although agreement does not rule out the possibility that common factors, such as confounding by indication, may be influencing results based on both sources.

Studies that have taken this approach and applied the same design protocol to more than one database have at times produced findings that closely agree, but have more often yielded inconsistent and even contradictory results. The largest of these studies systematically examined heterogeneity in relative risk estimates for 53 drug-outcome pairs across 10 US databases (all with more than 1.5 million patients) whilst holding the analytical method constant [8]. Around 30% of the drug-outcome pairs had effect estimates that ranged from a significantly decreased risk in some databases to a significantly increased risk in others; whilst only 13% were consistent in both direction and significance across all databases. However, there was wide variability between the datasets, which ranged from commercial insurance claims data to electronic health records, and from Medicare recipients to US veterans to privately insured citizens. Most other comparative studies have likewise been based on quite disparate databases, such as different countries [9-13], different geographical areas of the same country [10,11], different patient populations within a country [8], or different kinds of databases (for example, administrative claims data and electronic health records [8]).

These studies leave the reasons for the heterogeneity in results unclear: in particular the extent to which variability in results is due to differences in data recording and quality between databases, differences in demographics and health between the covered populations, or even a product of random processes and statistical artefacts. Untangling the factors driving heterogeneity of results is important for helping identify which data sources and results can be given credence and therefore usefully inform health decisions and policy [14].

To help address this issue, comparisons are useful that apply identical methods to two or more independent databases sampling from the same underlying patient population. By keeping the population and methods constant across databases, we can better determine the extent to which the database systems per se produce variability in the results. However, studies of this form are few and far between. Two replication studies using different UK PCDs reported closely corresponding results using different database sources [15-17], but these replications were conducted by research groups instrumental in the creation and maintenance of the comparator PCD and hence lacked independence. In a previous paper [1] we used the Clinical Practice

Research Datalink (CPRD) database [18] to conduct an exact and independent replication of a study originally undertaken in the QResearch database [19] on the impact of statins on survival in patients with coronary heart disease [20]. These databases have no practices in common and use data drawn from different practice electronic record systems (EMIS and VISION respectively). Reassuringly, our results using CPRD were in all main respects identical to those found with QResearch, in particular for the main outcome of overall risk of death associated with statins, which was lower by 55% in CPRD compared to 53% in QResearch.

In this paper we report on our second independent replication of a PCD study, comparing results derived from CPRD with results from a previously published study that used another PCD - the Doctors' Independent Network (DIN) [21] - that also does not overlap with CPRD, in either practices or record system. The original study by Shah and colleagues compared all-cause mortality in patients with a new diagnosis of solid cancer receiving beta-blockers with mortality in similar patients receiving alternative antihypertensive medications [22]. This represents a different clinical topic to those addressed by previous replications.

METHODS

The Doctors' Independent Network database (DIN) is an anonymised database drawing data from over 300 general practices using Torex software, covering over 3 million patients since 1989 [21]. There is no overlap between the practices in DIN and those in CPRD. An additional feature that makes DIN appealing for present purposes is that it is built around a quite different philosophy of how the medical record should be structured. CPRD records consultation notes as a sequence of discrete episodes, essentially unconnected, whereas DIN is based around the concept of the Problem Oriented Medical Record (POMR), which treats the medical record as a series of discrete but interconnected problems, with prescriptions linked to diagnoses under problem headings [23].

Data for 1998 for a subset of 142 DIN practices that passed data quality control checks for that year demonstrated very high comparability in age and gender structure to both CPRD and Office for National Statistics mid-year population estimates [21], although DIN practices are somewhat more likely to be located in Southern areas of the UK (Carey, personal correspondence). Prescription records are similar [23, 24] and good agreement has been reported for ischemic heart disease and hay fever prevalence [21] and the recording of thirty common childhood conditions [25].

As in our previous replication, we focused on studies of the effectiveness of medicinal interventions and after assessing the relevant studies that had been conducted in DIN, chose to replicate an investigation into the effects of beta-blocker treatment on cancer survival by Shah and colleagues [22]. This study concerned a quite different patient group and class of drug than our previous replication, and a small effect size in contrast to a large one. In addition, the topic under investigation was an incidental drug effect - suggested by earlier in vitro studies [26] - that sparked a great deal of medical community interest and related research activity, still ongoing [27]. The results of this activity have been very mixed and often contradictory, with some studies finding a protective effect for beta-blocker use in relation to mortality from breast cancer [28, 29] and others finding no effect or a modestly increased risk for various cancers, including lung, breast, and prostate [30] and substantially increased risk of developing more advanced colon cancer [31]. Interpretation of this variation in results is not simple as there are many differences between the studies, including the types of beta-blocker involved, which could influence the relationship to mortality [32].

Notwithstanding this complexity, for the purposes of this paper we are primarily concerned with the findings of the particular study by Shah and colleagues, whose DIN-based analysis produced some evidence for an

increase in all-cause mortality in cancer patients receiving beta-blockers. The size of the effect across the total sample was small, but sub-group analyses suggested that this reflected larger effects mostly confined to patients with pancreatic and prostate cancers and those on beta-blockers without additional blood pressure lowering medications. In their paper, Shah and colleagues acknowledge that they cannot easily explain these results, but conclude that their study does not support the hypothesis that beta- blockers improve survival for common cancers.

Using CPRD, we replicated the methods of Shah and colleagues as closely as possible, given the differences between the two databases. The methodological details provided in the published paper were not sufficient by themselves to allow a close replication to be conducted, and we therefore obtained additional details from the authors. We requested purely factual information about the methods used and did not share any of our analyses or results. All of the methods described below, including the study period, variable specifications and analytical procedures, are exact replications of those used in the original study, unless indicated otherwise.

We selected all practices in CPRD that provided up to standard (UTS) data (UTS is CPRD's designation for data meeting their internal quality standards) for the whole of the period from 1 January 1997 to 31 December 2006. Within these practices and period, we next identified all patients aged 40-85 with a first diagnosis of a solid tumour of the breast, lung, stomach, oesophagus, colon, renal system, prostate or ovary, and with at least two prescriptions of an anti-hypertensive drug (beta-blockers, ACE inhibitors, angiotensin receptor blockers, thiazides, calcium channel blockers, alpha-adrenoceptor blockers) in the year prior to diagnosis. We excluded patients with specific indications (coronary heart disease, heart failure, arrhythmias, stroke) or contra-indications (COPD, diabetes, asthma, renal disease) for antihypertensive medication that may impact on survival, prior to cancer diagnosis. Indications were determined using the Read code lists for the original study as provided to us by Shah and colleagues. We then classified the remaining patients according to exposure in the 1-year period prior to cancer diagnosis into three groups: (i) beta-blockers plus other blood-pressure lowering medications (BPLM); (ii) beta- blockers but no other BPLM; (iii) other BPLM only (controls). All Read codes used in the study are available on the ClinicalCodes repository at https://clinicalcodes.rss.mhs.man.ac.uk [33].

We extracted data for these patients from 1 year prior to cancer diagnosis, up until the end of 2007, or until the last recorded date for practices that stopped providing data before the end of 2007, giving a maximum possible length of follow-up post-diagnosis of 10 years. We intentionally made no attempt to 'improve' on the analysis conducted by Shah and colleagues as our specific aim was to determine whether the same results and conclusions would emerge from using identical methods on a different underlying dataset.

Analysis

The main outcome was all cause mortality, identified through a record of death in the CPRD. Patients who left their practice during the follow-up period were treated as censored observations in the analysis. Analysis used a Cox proportional hazards model with adjustment for patient age (below 55, 55-65, 66-75, 76 or older), gender, year of diagnosis, smoking status (current, ex-smoker, never smoked, not recorded, as recorded in the year prior to diagnosis), number of medications received in year prior to diagnosis, Regional Health Authority, and practice postcode Index of Multiple Deprivation [34]. The only measure not defined in the same way as by Shah and colleagues was deprivation, which was at the patient level in DIN but the practice in CPRD (see below).

Again following Shah and colleagues, we conducted an analysis for each cancer site separately, and then combined across sites using a DerSimonian-Laird random-effects meta-analysis. Analyses were undertaken to compare all patients receiving beta- blockers with the controls, and also for patients subdivided into those receiving, and those not receiving, additional BPLM. We further undertook analyses for non-selective beta-blockers only.

To make a direct comparison of the overall and cancer-specific treatment effect estimates (hazard ratios) from CPRD with those reported by Shah and colleagues for DIN, we took the natural log of each hazard ratio and estimated the standard error from the (logged) confidence interval limits, then ran a standard Wald test [35].

All analyses were performed using R version 3.0.2 [36]. In line with Shah and colleagues, we used an alpha level for statistical significance of 5% throughout.

Sensitivity analyses

We repeated the sensitivity analysis in the original study by excluding patients with less than 1-year survival after cancer diagnosis. The definitions of death and deprivation differed between databases and to assess sensitivity to this we repeated the analyses with the CPRD sample restricted to practices for which Office of National Statistics (ONS) official death dates and patient-level IMD scores were available (58% of practices and 60% of patients).

We observed notable differences between the cohorts regarding cancer site prevalence rates, area deprivation, year of diagnosis, and patient gender (table 1). Some of these differences, particularly year of diagnosis, are likely related to a considerable increase in the number of practices in the CPRD over the time of the study, compared to DIN (Supp table S1). To examine the sensitivity of our results to these differences we performed a sensitivity analysis on samples of the CPRD data that matched the make-up of the DIN cohort in key aspects.

To do this, we used an Iterative Proportional Fitting (IPF) [37] algorithm, a method for matching marginal distributions that does not assume independence between the matching variables. The matching variables were cancer site prevalence, year of diagnosis and area deprivation. The algorithm calculated selection probabilities (weights) for each patient in the CPRD data that were used to draw 10,000 weighted bootstrap samples (i.e. samples with replacement). Each sample was analysed and the results combined to obtain overall estimates of effect (the median hazard ratio) and 95% confidence intervals (2.5 and 97.5 percentiles). The IPF algorithm produced an excellent level of agreement on all three matching variables and also corrected the imbalance on gender but not smoking status (Supp table S2). On average, each bootstrap sample consisted of 1,352 patients on beta-blockers and 2,753 on other BPLMs. We also ran an analysis adjusting for the clustering of patients within practices, which Shah and colleagues performed but did not report as it made no difference to results (personal correspondence).

RESULTS

Comparison of patient cohorts

Table 1 compares the patient cohorts from CPRD and DIN on key measures. As expected from the greater number of practices in CPRD, the total sample was much larger (11302 versus 3462). Patients in the CPRD cohort were more likely to be male (55% v 47%), to be an ex-smoker (44% v 27%), to live in a more deprived

area (38% v 25% in the two most deprived quintiles) and to have been more recently diagnosed with cancer (69% v 51% since 2003). CPRD patients were also more likely to have a higher number of recorded medications (56% v 44% on 10 or more medications), though the rate of beta-blocker prescription in CPRD was a little lower (36% versus 41%). The breakdown of types of beta-blockers used was similar in both cohorts, with around three-quarters of patients on atenolol (based on their last prescription before cancer diagnosis): this minimises the risk that the comparison might be affected by differential associations between mortality and beta-blocker type [32].

There was a much higher rate of prostate cancer in the CPRD cohort (Table 2: 33% v 22%) but lower rates of ovarian and renal cancers – though absolute numbers of these were low in both cohorts. Rates for other types of cancer were all similar. The overall mortality rate was also considerably higher in CPRD (Table 2: 50% v 42%), though median lengths of follow-up were similar (29 v 30 months), as were survival rates 1 year post-diagnosis (78% v 74%). The disparity in overall mortality rates was not resolved by matching the CPRD and DIN samples (Table S2: 55% v 42%), nor was it resolved after further matching the samples on smoking status (51% v 42%), and hence cannot be attributed to cohort differences in cancer rates, year of diagnosis, or patient demographic factors.

Survival analysis (Table 3; figure 1)

Overall mortality

There was no difference in adjusted mortality rates between patients in the CPRD receiving beta-blockers (with or without other BPLM) and those on other BPLM only (overall hazard ratio=1.01, 95%CI 0.91 to 1.13). This compares to a small but statistically significant impact of beta-blockers on mortality in DIN (hazard ratio = 1.18, 95%CI 1.04 to 1.33). However, the Wald test directly comparing these hazard ratios was not statistically significant, although it did approach significance (p=0.063).

For the sub-set of patients on beta-blockers alone, we again found no significant impact on cancer mortality (HR=0.95, 95%CI 0.83 to 1.09), as opposed to a significant effect (HR=1.37, 95%CI 1.16 to 1.61) reported by Shah and colleagues. In this instance the comparison between studies was significant (p<0.001). For the remaining two subsets, of patients on beta-blockers plus other BPLM and patients on non-selective beta-blockers, the two studies returned similar, non-significant, results.

Mortality for individual cancer sites (Table 4)

Using CPRD, mortality rates for patients receiving beta-blockers were significantly higher for breast cancer (HR=1.19, 95%CI 1.03 to 1.37) and oesophageal cancer (HR=1.27, 95%CI 1.01 to 1.59) but significantly lower for patients with colon (HR=0.85, 95%CI 0.74 to 0.97) and renal cancer (HR 0.46, 95%CI 0.26 to 0.83), with no significant differences for other cancer sites. Using DIN, Shah and colleagues reported survival to be significantly poorer for patients with pancreas and prostate cancer, with no other differences. Thus for four of the nine cancer sites our CPRD study found a significant association of mortality with beta-blockers whereas the DIN study did not, and for two other sites this was reversed.

Direct comparison of the cancer site-specific hazard ratios from the two studies using Wald tests found no significant differences except for pancreatic cancer (p=0.023) and prostate cancer (p=0.016). For both cancers CPRD returned hazard ratios close to 1 whereas DIN produced much higher values. There was also significant heterogeneity of treatment effect across cancer sites in CPRD (p=0.004) in contrast to non-significant heterogeneity in DIN (p=0.41).

Results of sensitivity analysis

Sensitivity analysis using the subset of CPRD practices for which ONS mortality and patient-level IMD scores were available produced little change in the overall hazard ratio for death associated with beta-blockers (Table 4: HR=1.09, 95% CI 0.94 to 1.26). Hazard ratios for individual cancer sites likewise did not change greatly, although those for colon and renal cancers ceased to be statistically significant, at least partly due to the reduced sample.

Analysis of the CPRD sample(s) selected to match Shah and colleagues' DIN cohort resulted in an overall hazard ratio and 95% confidence interval identical to our primary analysis, and only small changes in the results for individual cancer sites, although the hazard ratios for oesophageal and renal cancers ceased to be statistically significant (table 4), as did the direct comparison of the CPRD and DIN hazard ratios for pancreatic cancers (p=0.069). Repeating our analyses adjusting for clustering of patients within practices and excluding patients who survived for less than a year made no substantive difference to any of the results (Table S3).

DISCUSSION

We conducted a fully independent, external, replication of a study based on one PCD using data from an alternative database. Our replication used the Clinical Practice Research Datalink (CPRD), a larger dataset than the Doctors' Independent Network (DIN), hence our total patient sample was more than three times the size of the original study. As far as we were able, we sampled from the same patient population and used identical methods to the original study, to minimise any sources of variation other than the database itself.

Our replication found no evidence for an association between beta-blocker use and cancer mortality, either in the full CPRD cohort or for patients on beta-blockers only - where the strongest effect was observed by Shah and colleagues using DIN. Results for individual cancer types also differed considerably, indicating an entirely different set of statistically significant cancer sites. However, most study differences disappeared under direct comparison of the hazard ratio estimates, with only the treatment effects for patients with pancreatic and prostate cancers and for those on beta-blockers alone remaining significantly different, for whom beta-blocker use was associated with mortality in DIN but not CPRD. Thus with these exceptions, all treatment effect estimates from the two studies agreed within the range of random variation. These results were unchanged in all essentials under sensitivity analyses using CPRD subsamples with linked ONS mortality and patient-level deprivation measures, and matched to DIN on cancer prevalence rates and other sample characteristics.

It is informative to compare both of these studies to a series of investigations by a group centred at Queen's University, Belfast, who also used CPRD to investigate the effects of beta-blocker usage on mortality from breast [27], colon [38] and prostate cancer [39], using a methodology that differed in a number of respects. No significant associations were found for any of these cancer sites, in contrast to both ourselves (breast, colon) and to Shah and colleagues (prostate). However, the confidence intervals reported by all three teams overlapped considerably - with the exception of prostate cancer from Shah and colleagues (table S4) – indicating that all treatment effect estimates were equivalent within statistical limits of accuracy.

Unresolved differences between CPRD and DIN after direct comparison were few and mainly low-level, but it is worthwhile to consider why these should remain. Results did show some sensitivity to database

differences in patient demographics, though not enough to explain all the discrepant results. The very different clinical computing systems used may have affected aspects of recorded care – possibly more than any practice or sample characteristic [40] - but the small number of study differences suggests that any overall impact was minimal, though an influence on specific data items and results is plausible. The much lower mortality rate in DIN suggests that death may have been defined differently, or recorded less reliably, but for this to explain the difference in effect estimates, the act of recording mortality in DIN would have to be associated with prescription of beta-blockers but not prescription of other BPLMs, and only for certain cancers but not others, which seems unlikely.

The unresolved differences might simply be statistical artefacts. Unmeasured confounding factors could vary in distribution between the datasets. Also, all the discrepant results concerned largely exploratory sub-group analyses done within a framework of multiple significance testing and arguably an alpha-level higher than 5% would be more appropriate: at alpha=1%, the only unresolved difference is for patients on beta-blockers alone. The QResearch replication study [20] likewise identified a number of within- and between-study discrepancies in subgroup analyses, relating to different statin compounds, and in all three of the Belfast group's studies, despite no overall associations sub-group analyses found significant relationships between cancer survival and from one to three specific beta-blocker compounds, though not always the same compound and not always in the same direction [27, 38, 39]. The large scale of many EHRs may encourage researchers to undertake multiple subgroup analyses without any firm hypotheses, and may also foster the idea that size alone offers some protection against incorrect inference, yet the rate at which inconsistent results occur in EHR-based studies strongly suggests that issues of multiple testing, "fishing" for results, and spurious significance apply as much to these data sources as they do to much smaller datasets; possibly even more so given the potential for bias from residual and uncontrolled confounding.

The results of our study therefore present a somewhat complex picture: examined separately and purely in terms of statistical significance, the CPRD- and DIN-based studies provided rather different pictures of the risks of beta-blockers overall and in relation to different cancer types. However, when directly compared and excepting a few specific subgroup analyses, estimates of treatment effect did not differ statistically. The overall clinical conclusion was also the same: that beta-blockers offer no survival advantage to cancer patients. The present study in combination with our previous independent replication study and other non-independent replications, therefore constitutes a growing body of evidence that the application of identical analytical methods to a variety of different UK PCDs tends to yield treatment effect estimates that are in the main comparable within statistical limits of accuracy.

Limitations

Differences between the DIN and CPRD databases meant that while we were able to exactly replicate the great majority of the components of the original study, there were a few exceptions. The datasets may have differed in their definitions of all-cause mortality, as each use their own bespoke algorithm. For area deprivation, Shah and colleagues used 2004 IMD scores in national quintiles based on each patient's postcode. Equivalent scores were only available to us for a subset of CPRD, so instead we used 2004 practice-postcode IMD scores, obtained for all practices from the CPRD organisation as a linked dataset. We tested for the impact of these factors by running a sensitivity analysis using the subset of CPRD patients for which linked ONS data on the date of death and residential IMD 2004 scores were available. In all other respects, this study replicated the original with respect to the population and variable definitions and methods of analysis.

The overall raw mortality rate in our CPRD cohort was substantially higher than in the DIN cohort and a much higher proportion of CPRD cancers were of the prostate. Patients in the CPRD cohort were also likely to have been diagnosed more recently, to live in areas of higher deprivation and to be male. However, analysis of subsets of the CPRD cohort matched to DIN did not account for the difference in overall mortality rates, nor did it substantially alter our findings. Neither the complete details of how Shah and colleagues defined mortality nor of the CPRD mortality algorithm, were available to us, thus our ability to uncover the reasons for these different mortality rates was limited.

We intentionally did not try to improve on the analysis methods used by Shah and colleagues, even though these have received some criticism [32, 38], since for our purposes it was important to keep the analysis methods constant. Criticisms include: not linking to cancer registries; lack of control for stage of disease or treatment; not differentiating beta-blocker use prior- and post- cancer diagnosis; and use of patients on other antihypertensives as the comparator. Most of these criticisms were in fact been discussed by Shah and colleagues in their paper and justified there as part of the methods. Importantly, the Belfast group's CPRD-based studies took account of most of these issues yet yielded effect estimates very similar to our own.

Conclusion

This replication of one UK PCD-based study in a second completely independent PCD using the same methods and sampling the same population has revealed a complex pattern of similarities and differences in both the make-up of the patient cohorts and in the findings from analysis. However, when directly compared, with the exception of a few specific subgroup results estimates of treatment effect did not differ statistically. The present study therefore adds to previous replication work in finding that when analysed to a common protocol, different UK PCDs produce treatment effect estimates with reasonably close - though not exact - agreement. However, these particular databases have been shown to possess high degrees of validity for most key data items and to provide good coverage of the UK primary care patient population, and it would be a mistake to assume the same applies to other less well-validated or to subpopulation-specific EHR datasets. Our results also suggest that where relatively small effects and subgroup results are concerned, even these well-validated databases do not guarantee generalizable results and great care must be taken in drawing any firm conclusions. In all cases, confirmatory studies using at least one other independent data source are strongly recommended.

REFERENCES

1. Reeves D, Springate DA, Ashcroft DM, et al. Can analyses of electronic patient records be independently and externally validated? The effect of statins on the mortality of patients with Ischaemic heart disease: a cohort study with nested case-control analysis. *BMJ Open* 2014;4:e004952

2. Herrett E, Thomas SL, Schoonen WM, et al. Validation and validity of diagnoses in the General Practice Research Database: a systematic review. *Br J Clin Pharmacol* 2010;69:4-14.

3. Khan NF, Harrison SE, Rose PW. Validity of diagnostic coding within the General Practice Research database: a systematic review. *BJGP* 2010;e128-e136.

4. Jordan K, Porcheret M, Croft P. Quality of morbidity coding in general practice computerized medical records: a systematic review. *Family Practice* 2004;21:396-412.

5. Marston L, Carpenter JR, Walters KR, et al. Issues in multiple imputation of missing data for large general practice clinical databases. *Pharmacoepidemiol Drug Saf* 2010;19:618-626.

6. Delaney JA, Daskalopoulou SS, Brophy JM, et al. Lifestyle variables and the risk of myocardial infarction in the General Practice Research Database. *BMC Cardiovascular Disorders* 2007;7:38.

7. Collins SC, Altman DG. An independent and external validation of QRISK2 cardiovascular disease risk score: a prospective open cohort study. BMJ 2010 340:c2442.

8. Madigan D, Ryan PB, Schuemie M, et al. Evaluating the Impact of Database Heterogeneity on Observational Study Results. *Am J Epidemiol* 2013;178:645-651.

9. Wijnans L, Lecomte C, de Vries C, et al. The incidence of narcolepsy in Europe: Before, during, and after the influenza A(H1N1)pdm09 pandemic and vaccination campaigns. *Vaccine* 2013;31:1246-1254.

10. Dormuth CR, Hemmelgarn BR, Paterson JM, et al. Use of high potency statins and rates of admission for acute kidney injury: multicenter, retrospective observational analysis of administrative databases. *BMJ* 2013;346:f880

11. Filion KB, Chateau D, Targownik LE, et al. Proton pump inhibitors and the risk of hospitalisation for community-acquired pneumonia: replicated cohort studies with meta-analysis. *Gut* 2014;63:552-8.

12. Valkhoff VE, van Soest EM, Mazzaglia G, et al. Adherence to gastroprotection during cyclooxygenase 2 inhibitor treatment and the risk of upper gastrointestinal tract events: a population-based study. *Arthritis Rheum* 2012;64:2792-2802.

13. Andrews N, Stoweb J, Miller E, et al. A collaborative approach to investigating the risk of thrombocytopenic purpura after measles–mumps–rubella vaccination in England and Denmark. *Vaccine* 2012;30:3042-3046.

14. Bell C, Chakravarty A, Gruber S, et al. Characteristics of study design and elements that may contribute to the success of electronic safety monitoring systems. *Pharmacoepidemiol Drug Saf* 2014;DOI:10.1002/pds.3712

15. Bremner SA, Carey IM, DeWilde S, et al. Early-life exposure to antibacterials and the subsequent development of hayfever in childhood in the UK: case-control studies using the General Practice Research Database and the Doctors' Independent Network. *Clin Exp Allergy* 2003;33:1518-25.

16. Bremner SA, Carey IM, DeWilde S, et al. Vaccinations, infections and antibacterials in the first grass pollen season of life and risk of later hayfever. *Clin Exp Allergy* 2007;37:512-517.

17. Vinogradova Y, Coupland C, Hippisley-Cox J. Exposure to bisphosphonates and risk of gastrointestinal cancers: series of nested case-control studies with QResearch and CPRD data. *BMJ* 2013;346:f114 doi: 10.1136/bmj.f114.

18. Clinical Practice Research Datalink (CPRD). Available: http://www.cprd.com/home. Date accessed 20 March 2014.

19. QResearch. Available: http://www.qresearch.org. Date accessed 20 March 2014

20. Hippisley-Cox J, Coupland C. Effect of statins on the mortality of patients with ischaemic heart disease: population based cohort study with nested case-control analysis. *Heart* 2006;92:752-758.

21. Carey IM, Cook DG, De Wilde S, et al. Developing a large electronic primary care database (Doctors' Independent Network) for Research. *Int J Med Inform* 2004;73:443-453.

22. Shah SM, Carey IM, Owen CG, et al. Does b-adrenoceptor blocker therapy improve cancer survival? Findings from a population-based retrospective cohort study. *Br J Clin Pharmacol* 2011;72:157–161.

23. Carey IM, Cook DG, De Wilde S, et al. Implications of the problem orientated medical record (POMR) for research using electronic GP databases: a comparison of the Doctors Independent Network Database (DIN) and the General Practice Research Database (GPRD). *BMC Fam Pract* 2003;4: 14.

24. De Wilde S, Carey IM, Bremner SA, et al. Evolution of statin prescribing 1994-2001: a case of agism but not of sexism? *Heart* 2003;89:417-421.

25. De Wilde S, Carey IM, Bremner S. A comparison of the recording of 30 common childhood conditions in the Doctor's Independent Network and General Practice Research Databases. *Health Stat Q* 2004;22:21–31.

26. Chakroborty D, Sarkar C, Basu B, et al. Catecholamines regulate tumor angiogenesis. *Cancer Res* 2009;69:3727–30.

27. Cardwell CR, Coleman HG, Murray LJ, et al. Beta-blocker usage and breast cancer survival: a nested case-control study within a UK Clinical Practice Research Datalink cohort. *Int J Epidemiol* 2013;42:1852-61 doi: 10.1093/ije/dyt196.

28. Ganz PA, Habel LA, Weltzien EK, et al. Examining the influence of beta blockers and ACE inhibitors on the risk for breast cancer recurrence: results from the LACE cohort. *Breast Cancer Res Treat* 2011;129:549–556.

29. Melhem-Bertrandt A, Chavez-Macgregor M, Lei X, et al. Beta-blocker use is associated with improved relapse-free survival in patients with triple-negative breast cancer. *J Clin Oncol* 2011;29:2645–2652.

30. Friedman GD, Udaltsova N, Habel LA. Norepinephrine antagonists and cancer risk. *Int J Cancer* 2011;128:737–738.

31. Jansen L, Below J, Chang-Claude J, et al. Beta blocker use and colorectal cancer risk: population-based case-control study. *Cancer* 2012;118:3911-3919.

32. Barron TI, Sharp L, Visvanathan K. Beta-adrenergic blocking drugs in breast cancer: A perspective review. *Ther Adv Med Oncol* 2012;4:113–125 doi: 10.1177/1758834012439738

33. Springate DA, Kontopantelis E, Ashcroft DM, et al. ClinicalCodes: An Online Clinical Codes Repository to Improve the Validity and Reproducibility of Research Using Electronic Medical Records. *PLoS ONE* 2014;9:e99825 doi:10.1371/journal.pone.0099825

34. Available: http://data.gov.uk/dataset/imd_2004. Date accessed 20 October 2014.

35. Tierney JF, Stewart LA, Ghersi D, et al. Practical methods for incorporating summary time-to-event data into meta-analysis. *Trials* 2007;8:16

36. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing 2013. Available: http://R-project.org. Date accessed 20 November 2014

37. Speed TP. Iterative Proportional Fitting. *Encyclopedia of Biostatistics* 2005;4:doi: 10.1002/0470011815.b2a10027

38. Hicks BM, Murray LJ, Powe DG, et al. β-Blocker usage and colorectal cancer mortality: a nested case-control study in the UK Clinical Practice Research Datalink cohort. *Ann Oncol* 2013;24:3100-6 doi: 10.1093/annonc/mdt381.

39. Cardwell CR, Coleman HG, Murray LJ, et al. Beta-blocker usage and prostate cancer survival: a nested case-control study in the UK Clinical Practice Research Datalink cohort. *Cancer Epidemiol* 2014;38:279-85 doi: 10.1016/j.canep.2014.03.011.

40. Kontopantelis E, Buchan I, Reeves D, et al. Relationship between quality of care and choice of clinical computing system: retrospective analysis of family practice performance under the UK's quality and outcomes framework. *BMJ Open* 2013;3:e003190 doi:10.1136/bmjopen-2013-003190

Acknowledgements
This study is based on data from the Clinical Practice Research Datalink (CPRD) obtained under licence from the UK Medicines and Healthcare products Regulatory Agency. However, the interpretation and conclusions contained in this paper are those of the authors alone. We are grateful to Dr Iain M. Carey from the Division of Population Health Sciences and Education, St George's University of London, for his comprehensive responses to our queries about the original DIN study by himself, Shah and colleagues and about DIN itself.

Contributors
DR, EK, and RR developed the original idea for the study and DS, DA and TD contributed to the study design. DR supervised all aspects of the study's execution. DS helped to plan the analysis and undertook the primary analysis. DR and DS wrote the first draft of the paper. All authors critically reviewed the paper and approved the submitted version. DS had full access to all the data in the study and he takes responsibility for the integrity of the data and the accuracy of the data analysis.

Funding
This work was funded by the National Institute for Health Research School for Primary Care Research (project number 142). The funders had no role in the conducting of the study or in the decision to submit the paper for publication. The views expressed are those of the authors and not necessarily those of the NHS, the National Institute for Health Research or the Department of Health.

Competing interests
All authors have completed the Unified Competing Interest form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare that DR was partly funded and DS fully funded by an NIHR School for Primary Care Research grant to undertake this study and EK was partly supported by a NIHR School for Primary Care Research fellowship in primary health care but that all other authors received no support from any additional organisation for the submitted work and had no relationships with any additional organisation that might have an interest in the submitted work in the previous 3 years. For all authors, their spouses, partners, or children have no financial relationships that may be relevant to the submitted work; and no non-financial interests that may be relevant to the submitted work.

Ethics approval
The study protocol was approved by the independent scientific advisory committee (ISAC) for CPRD research (reference number: 12_149R).

Data sharing
Clinical Practice Research Datalink data cannot be shared due to licencing restrictions. All the code-lists used in the analysis of CPRD in this study are available at https://clinicalcodes.rss.mhs. man.ac.uk/.The full R code used for the analysis of CPRD is available from the authors.

Table 1. Comparison of CPRD and DIN patient cohorts (%, (n))

Variable		CPRD (n=11302)	DIN (n=3462)
Age at diagnosis (years)	18-55	5.3% (602)	NA
	56-65	23.3% (2631)	NA
	66-75	37.4% (4228)	NA
	75 and above	33.9% (3841)	NA
Gender	Male	55.3% (6247)	47.4% (1641)
	Female	44.7% (5055)	52.6% (1821)
Smoking	Current	14.7% (1665)	19.1% (661)
	ex-smoker	43.9% (4962)	27.2% (941)
	Never smoked	34.2% (3864)	51.8% (1792)
	Missing	7.2% (811)	1.9% (68)
Deprivation (IMD 2004 Quintiles)[a]	1 (most deprived)	17.8% (2012)	9.6% (333)
	2	19.9% (2253)	14.9% (517)
	3	21% (2375)	19.2% (664)
	4	23.1% (2613)	22.0% (760)
	5 (least deprived)	18.1% (2049)	26.4% (915)
	Missing	0% (0)	7.9% (273)
Year of diagnosis	1997-8	5.8% (658)	12.1% (420)
	1999-00	8.8% (996)	15.8% (546)
	2001-2	16.4% (1856)	20.6% (714)
	2003-4	29.2% (3303)	25.1% (870)
	2005-6	39.7% (4489)	26.3% (912)
N of medications	0-4	14.9% (1681)	16.9% (586)
	5-9	29.4% (3319)	38.1% (1318)
	10-14	28.2% (3188)	24.4% (845)
	15-19	8.5% (956)	9.6% (332)
	20 and above	19.1% (2158)	7.8% (269)
	Missing	0% (0)	3.2 (112)
Prescribed b-blocker	No	64.3% (7272)	59.4% (2057)
	Yes	35.7% (4030)	40.6% (1405)
Type of beta-blocker	Atenolol	73.0% (2943)	75.2% (1057)
	Propranolol	11.0% (443)	12.8% (180)
	Other beta-blocker	16.0% (644)	12.0% (168)

NA = Not Available

[a]Based on patient postcode for DIN and practice postcode for CPRD

Table 2. Comparison of CPRD and DIN patient cohorts by exposure to blood pressure lowering medication (BPLM) in the year prior to cancer diagnosis

	All patients				No beta-blockers (controls: other BPLM but no beta-blockers)				Multiple hypertensives (beta-blockers plus other BPLM)				Beta-blockers only (beta-blockers but no other BPLM)			
	CPRD		DIN		CPRD		DIN		CPRD		DIN		CPRD		DIN	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
All patients	11302	100	3462	100	7272	64.3	2056	59.4	2832	25.10	864	24.4	1198	10.6	542	15.7
Deaths	5754	50.1	1441	41.6	3748	51.6	846	41.2	1459	51.5	350	40.5	547	45.7	245	45.2
Alive at 1-year follow-up	8763	77.5	2576	74.4	5607	77.1	1541	75.0	2194	77.5	630	72.9	962	80.3	405	74.7
On non-selective beta-blocker	685	6.06	239	7.5	NA	NA	NA	NA	359	12.7	71	8.2	326	27.2	167	30.8
Cancer sites																
Breast	2943	26.0	984	28.4	1746	24.0	554	26.9	794	28.0	240	27.8	403	33.6	194	35.8
Colon	1799	15.9	619	17.9	1104	15.2	354	17.2	476	16.8	162	18.7	219	18.3	103	18.9
Lung	1326	11.7	436	12.6	913	12.6	277	13.5	307	10.9	105	12.1	106	8.8	54	9.9
Oesophagus	434	3.8	159	4.6	257	3.5	95	4.6	116	4.1	44	5.1	61	5.1	20	3.7
Ovarian	203	1.8	148	4.3	124	1.7	76	3.7	45	1.6	43	5.0	34	2.8	29	5.4
Pancreas	376	3.3	140	4.0	222	3.1	83	4.0	111	4.0	34	3.9	43	3.6	23	4.2
Prostate	3748	33.2	759	21.9	2604	35.8	500	24.3	856	30.2	182	21.0	288	24.0	77	14.1
Renal	141	1.2	124	3.6	81	1.1	69	3.4	48	1.7	33	3.8	12	1.0	22	4.0
Stomach	332	2.9c	93	2.7	221	3.0	52	2.5	79	2.8	21	2.4	32	2.7	20	3.7

NA = Not Applicable

Table 3. Comparison of patients using beta-blockers versus controls from the CPRD and DIN studies: pooled hazard ratios (95% CI)) from meta-analyses of cancer site-specific results

Comparison (versus controls)	CPRD	DIN	Wald *p*-value
All patients using beta-blockers	1.01 (0.91, 1.13)	1.18 (1.04 to 1.33)**	0.06
Patients using beta-blockers only	0.94 (0.82, 1.07)	1.37 (1.16, 1.61)***	<0.001***
Beta-blockers and other BPLM	1.06 (0.95, 1.19)	1.11 (0.91, 1.34)	0.69
Non-selective beta-blockers only	0.96 (0.8, 1.15)	1.21 (0.94, 1.55)	0.14

**p<0.01; ***p<0.001

Table 4. Cancer site-specific and pooled hazard ratios (95% CI)) from CPRD and DIN studies for all patients using beta-blockers versus controls

Cancer site	CPRD full cohort	CPRD sensitivity analysis 1: ONS mortality and & deprivation	CPRD sensitivity analysis 2: matched CPRD and DIN samples	DIN primary analysis	Wald *p*-value: CPRD full cohort v DIN (CPRD matched sample v DIN)
Sample size (on beta-blockers; controls)	4030; 7272	2528; 4514	4105[a]; 7197[a]	1406; 2056	-
Cancer sites					
Breast	1.19 (1.03, 1.37)*	1.28 (1.07, 1.32)***	1.24 (1.09, 1.43)**	1.09 (0.80, 1.49)	0.62 (0.46)
Colon	0.85 (0.74, 0.97)*	0.9 (0.76, 1.07)	0.87 (0.76, 1.0)*	1.00 (0.77, 1.30)	0.28 (0.36)
Lung	1.04 (0.91, 1.19)	0.97 (0.82, 1.15)	1.03 (0.88, 1.21)	1.12 (0.89, 1.41)	0.59 (0.56)
Oesophagus	1.27 (1.01, 1.59)*	1.65 (1.22, 2.24)***	1.19 (0.93, 1.52)	1.05 (0.69, 1.60)	0.44 (0.61)
Ovarian	1.05 (0.74, 1.5)	1.5 (0.89, 2.52)	1.02 (0.77, 1.33)	1.14 (0.63, 2.06)	0.82 (0.74)
Pancreas	0.94 (0.74, 1.21)	0.82 (0.6, 1.11)	1.04 (0.75, 1.43)	1.88 (1.09, 3.25)*	0.023* (0.069)
Prostate	1.03 (0.92, 1.15)	1.02 (0.88, 1.17)	0.94 (0.82, 1.08)	1.54 (1.13, 2.09)**	0.016* (0.004)**
Renal	0.46 (0.26, 0.83)**	0.6 (0.28, 1.27)	0.67 (0.41, 1.03)	1.14 (0.52, 2.52)	0.069 (0.25)
Stomach	1.03 (0.78, 1.36)	1.34 (0.93, 1.93)	1.01 (0.68, 1.48)	1.44 (0.76, 2.74)	0.35 (0.35)
All patients using beta-blockers	1.01 (0.91, 1.13)	1.09 (0.94, 1.26)	1.01 (0.91, 1.13)	1.18 (1.04, 1.33)**	0.063 (0.063)

[a]Median across bootstrap samples

* p<0.05; ** p<0.01; ***p<0.001

Figure 1. Hazard ratios of survival for patients prescribed beta- blocker therapy compared to patients prescribed other BPLM

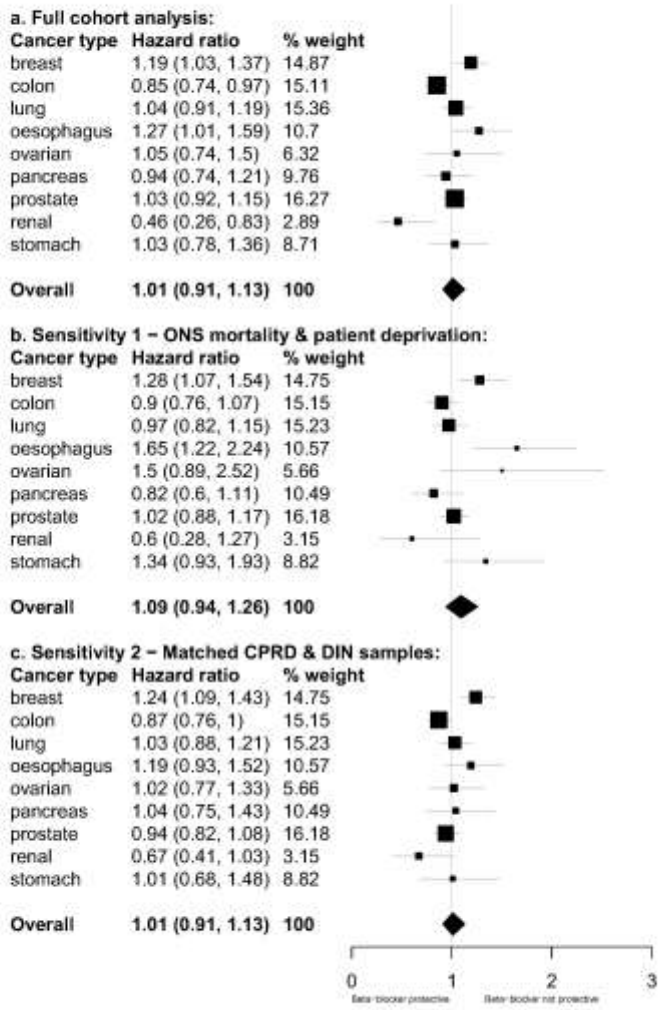

Figure 1. Hazard ratios of mortality for patients prescribed beta- blocker therapy compared to patients prescribed other BPLM
203x269mm (300 x 300 DPI)

Table S1: Numbers of practices and newly diagnosed patients in each cohort (practices with one or more eligible patients only)

Year	CPRD		DIN	
	N of practices	N of patients	N of practices	N of patients
1997	155	303	88	192
1998	172	355	100	228
1999	200	398	99	253
2000	254	598	119	293
2001	303	808	119	360
2002	350	1048	124	354
2003	425	1427	140	426
2004	495	1876	148	444
2005	510	2090	143	474
2006	526	2399	140	438

Table S2: Comparison of CPRD and DIN patient cohorts, matched samples (%, (n))

Variable		**CPRD matched samples[a] (n=11,302)**	**DIN (n=3462)**
Cancer site	Breast	28.7% (3244)	28.4% (984)
	Colon	18.1% (2042)	17.9% (619)
	Lung	12.7% (1437)	12.6% (436)
	Oesophagus	4.5 (514)	4.6% (159)
	Ovarian	3.8 (427)	4.3% (148)
	Pancreas	4.0 (455)	4.0% (140)
	Prostate	22.1 (2501)	21.9% (759)
	Renal	3.3 (372)	3.6% (124)
	Stomach	2.7 (307)	2.7% (93)
Year of diagnosis	1997-8	11.7% (1321)	12.1% (420)
	1999-00	15.6% (1765)	15.8% (546)
	2001-2	20.7% (2340)	20.6% (714)
	2003-4	25.4% (2868)	25.1% (870)
	2005-6	26.6% (3006)	26.3% (912)
Deprivation (IMD 2004 Quintiles)[b]	1 (most deprived)	10.4% (1175)	10.4% (333)
	2	16.2% (1831)	16.2% (517)
	3	20.8% (2340)	20.8% (664)
	4	23.9% (2871)	23.9% (760)
	5 (least deprived)	28.7% (3006)	28.7% (915)
Gender	Male	47.7% (5393)	47.4% (1641)
	Female	52.3% (5909)	52.6% (1821)
Smoking	Current	15.1% (1711)	19% (661)
	Ex-smoker	45.8% (5173)	27% (941)
	Never smoked	29.6% (3343)	52% (1792)
	Not recorded	9.5% (1075)	2% (68)
Prescribed beta-blocker	No	63.7% (7197)	59.4% (2057)
	Yes	36.3% (4105)	40.6% (1405)
Died	No	44.9% (5082)	58.4% (2021)
	Yes	55.1% (6220)	41.6% (1441)

[a]Counts and %'s for CPRD are medians across all bootstrapped samples

[b]Based on patient postcode for DIN, and practice postcode for CPRD

Table S3: Cancer site-specific and pooled hazard ratios (95% CI)) from CPRD for all patients using beta-blockers versus controls: additional sensitivity analyses

	CPRD sensitivity analysis 3: Clustering of patients by practice	CPRD sensitivity analysis 4: excluding patients who survived for < 1 year
Sample size (on beta-blockers; controls)	4030; 7272	3156; 5607
Cancer sites		
breast	1.19 (1.03, 1.36)*	1.22 (1.05, 1.43)*
colon	0.85 (0.73, 0.98)*	0.91 (0.77, 1.08)
lung	1.04 (0.89, 1.21)	0.97 (0.75, 1.25)
oesophagus	1.27 (0.99, 1.62)	0.99 (0.64, 1.52)
ovarian	1.05 (0.73, 1.51)	1.41 (0.87, 2.29)
pancreas	0.94 (0.71, 1.25)	1.81 (0.84, 3.92)
prostate	1.03 (0.92, 1.15)	1.02 (0.9, 1.16)
renal	0.46 (0.25, 0.85)*	0.63 (0.29, 1.34)
stomach	1.03 (0.77, 1.37)	1.17 (0.67, 2.06)
All patients using beta-blockers	1.01 (0.91, 1.13)	1.05 (0.94, 1.18)

* $p<0.05$

Table S4: Summary of results for breast, colon and prostate cancer for the current study, Shah et al and the Belfast Group (BG)

Cancer type	Current (CPRD): cohort analysis, all-cause mortality (Hazard Ratio)	Shah et al (DIN): cohort analysis, all-cause mortality (Hazard Ratio)	BG[1] (CPRD): cohort analysis, cancer specific mortality (Hazard Ratio)	BG[1] (CPRD): case-control analysis, all-cause mortality (Odds Ratio)
Breast	1.19 (1.03, 1.37)*	1.09 (0.80, 1.49)	0.93 (0.83, 1.06)	1.03 (0.92, 1.16)
Colon	0.85 (0.74, 0.97)*	1.00 (0.77, 1.30)	0.90 (0.79, 1.01)	0.88 (0.77, 1.0)
Prostate	1.03 (0.92, 1.15)	1.54 (1.13, 2.09)**	0.98 (0.86, 1.12)	1.05 (0.94, 1.17)

[1]The BG reported hazard ratios from cohort analysis for cancer-specific death only, but odds-ratios from case-control analysis for all-cause mortality. We therefore include both sets of results in the table.

*$p<0.05$; **$p<0.01$

STROBE 2007 (v4) checklist of items to be included in reports of observational studies in epidemiology*

Checklist for cohort, case-control, and cross-sectional studies (combined)

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	3-4
Objectives	3	State specific objectives, including any pre-specified hypotheses	3-4
Methods			
Study design	4	Present key elements of study design early in the paper	4-5
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	5
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	5 NA NA
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case	6 NA
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	5-6
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	5
Bias	9	Describe any efforts to address potential sources of bias	5-6
Study size	10	Explain how the study size was arrived at	5
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	5-6
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	5-6
		(*b*) Describe any methods used to examine subgroups and interactions	5
		(*c*) Explain how missing data were addressed	5
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed	5

		Cross-sectional study—If applicable, describe analytical methods taking account of sampling strategy	
		(*e*) Describe any sensitivity analyses	5-6
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	6
		(b) Give reasons for non-participation at each stage	NA
		(c) Consider use of a flow diagram	See cover letter
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	6 and table 1
		(b) Indicate number of participants with missing data for each variable of interest	Table 1
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	6
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	6, table 2 and figure 1
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	NA
		Cross-sectional study—Report numbers of outcome events or summary measures	NA
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	6-7, tables 3 and 4
		(*b*) Report category boundaries when continuous variables were categorized	5
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	NA
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	6-7
Discussion			
Key results	18	Summarise key results with reference to study objectives	7
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	8-9
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	9
Generalisability	21	Discuss the generalisability (external validity) of the study results	8-9
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	17

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

Can analyses of electronic patient records be independently and externally validated? Study 2: the Effect of Beta-Adrenoceptor Blocker Therapy on Cancer Survival; a Retrospective Cohort Study

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2014-007299.R1
Article Type:	Research
Date Submitted by the Author:	12-Mar-2015
Complete List of Authors:	Springate, David; University of Manchester, Centre for Primary Care Ashcroft, Darren; University of Manchester, Centre for Pharmacoepidemiology Kontopantelis, Evangelos; University of Manchester, Institute of Population Health Doran, Tim; University of York, Health Sciences Ryan, Ronan; University of Birmingham, Reeves, David; University of Manchester, Institute of Population Health; Centre for Biostatistics
Primary Subject Heading:	General practice / Family practice
Secondary Subject Heading:	Research methods, Evidence based practice, Health informatics, Oncology, Pharmacology and therapeutics
Keywords:	PRIMARY CARE, Health informatics < BIOTECHNOLOGY & BIOINFORMATICS, ONCOLOGY, STATISTICS & RESEARCH METHODS

Can analyses of electronic patient records be independently and externally validated? Study 2: the Effect of Beta-Adrenoceptor Blocker Therapy on Cancer Survival; a Retrospective Cohort Study

Corresponding author
David Reeves
NIHR School for Primary Care Research, Centre for Primary Care, Institute of Population Health, University of Manchester, Manchester M13 9PL, UK
Email: david.reeves@manchester.ac.uk
Tel: 0161 275 3536

Authors in order
David A Springate, *research fellow*1 3, Darren M Ashcroft, *professor*4, Evangelos Kontopantelis, *senior research fellow*1 2, Tim Doran, *professor5*, Ronan Ryan *research fellow6*, David Reeves, *reader*1 3

1 NIHR School for Primary Care Research, Centre for Primary Care, Institute of Population Health, University of Manchester
2 Centre for Health Informatics, Institute of Population Health, University of Manchester
3 Centre for Biostatistics, Institute of Population Health, University of Manchester
4 Centre for Pharmacoepidemiology and Drug Safety, Manchester Pharmacy School, University of Manchester
5 Department of Health Sciences, University of York
6 Primary Care Clinical Sciences, School of Health and Population Sciences, University of Birmingham

Keywords
Validity of electronic health records; primary care; health informatics; drug effectiveness; cohort study

Word count
4759

Abstract

Objectives To conduct a fully independent and external validation of a research study based on one electronic health record database, using a different database sampling the same population.

Design Retrospective cohort analysis of beta-blocker therapy and all-cause mortality in cancer patients.

Setting Two UK national primary care databases (PCDs): the Clinical Practice Research Datalink (CPRD) and Doctors' Independent Network (DIN).

Participants CPRD data for 11302 cancer patients compared to published results from DIN for 3462 patients. Study period January 1997 to December 2006.

Primary and secondary outcome measures All-cause mortality: overall; by treatment subgroup (beta-blockers only, beta-blockers plus other blood pressure lowering medicines (BPLM), other BPLMs only); and by cancer-site.

Results Using CPRD, beta-blocker use was not associated with mortality (HR=1.03, 95%CI 0.93-1.14, vs patients prescribed other BPLMs only), but in DIN beta-blocker users had significantly higher mortality (HR=1.18, 95%CI 1.04 to 1.33). However, these hazard ratios were not statistically different (p=0.063), but did differ for patients on beta-blockers alone (CPRD=0.94, 95%CI 0.82-1.07; DIN=1.37, 95%CI 1.16-1.61; p<0.001). Results for nine individual cancer sites differed by study, but only significantly for prostate and pancreas cancers. Results were robust under sensitivity analyses, but we could not be certain that mortality was identically defined in both databases.

Conclusions We found a complex pattern of similarities and differences between databases. Our finding that overall treatment effect estimates were not statistically different, adds to a growing body of evidence that different UK PCDs produce effect estimates comparable within statistical tolerance. However, some subgroup effects differed significantly and individually the two studies lead to different conclusions regarding the safety of beta-blockers for cancer patients. Single studies based on internally well-validated databases therefore do not guarantee generalisable results, especially for subgroups. In all cases, confirmatory studies using at least one other independent data source are strongly recommended.

Strengths and limitations of this study

- Drug effectiveness studies applying the same analysis protocol to different electronic health record (EHR) databases have typically compared EHRs covering different patient populations or replications have not been independently conducted. This paper reports on a fully independent validation of a published EHR-based study using a different EHR database sampling from the same underlying population.

- Despite purporting to cover the same general UK population, there were some notable demographic and clinical differences between the CPRD and DIN cancer cohorts. Sensitivity analysis indicated that these had only a minimal effect on treatment effect estimates, but we were unable to account for a difference in mortality rates between the cohorts.

- The present study adds to evidence from our previous independent replication study and other non-independent replications, that the application of identical analytical methods to a variety of different UK primary care databases produces treatment effect estimates that are in most respects comparable. Nevertheless, we also find that single studies, even when based on these well-validated data sources, do not guarantee generalizable results.

INTRODUCTION

Large-scale electronic health record databases (EHRs) are widely regarded as an important new tool for medical research. The major UK "primary care databases" (PCDs) are some of the largest and most detailed sources of electronic patient data available, holding detailed long-term clinical data for many millions of patients. Researchers are increasingly using these resources [1], which provide a means for researching questions in primary care that cannot feasibly be addressed by other means, including unintended consequences of drug interventions, where ethical considerations, the required numbers of patients, or length of follow-up can make an RCT impractical.

Concerns remain, however, about the validity of studies based on such data, including uncertainties about data quality, data completeness, and the potential for bias due to both measured and unobserved confounders. Most work on EHR validity has focused on the accuracy or completeness of the individually recorded data values, such as consultation recording [2], disease diagnoses [3,4] and risk factors [5-7]. Another approach to testing the validity of EHR-based studies is to compare the results to those obtained from equivalent investigations conducted on other, independent, datasets. Agreement of results helps to reassure that the findings do not depend upon the source of the data, although agreement does not rule out the possibility that common factors, such as confounding by indication, may be influencing results based on both sources.

Studies that have taken this approach and applied the same design protocol to more than one database have at times produced findings that closely agree, but have more often yielded inconsistent and even contradictory results. The largest of these studies systematically examined heterogeneity in relative risk estimates for 53 drug-outcome pairs across 10 US databases (all with more than 1.5 million patients) whilst holding the analytical method constant [8]. Around 30% of the drug-outcome pairs had effect estimates that ranged from a significantly decreased risk in some databases to a significantly increased risk in others; whilst only 13% were consistent in both direction and significance across all databases. However, there was wide variability between the datasets, which ranged from commercial insurance claims data to electronic health records, and from Medicare recipients to US veterans to privately insured citizens. Most other comparative studies have likewise been based on quite disparate databases, such as different countries [9-13], different geographical areas of the same country [10,11], different patient populations within a country [8], or different kinds of databases (for example, administrative claims data and electronic health records [8]).

These studies leave the reasons for the heterogeneity in results unclear: in particular the extent to which variability in results is due to differences in data recording and quality between databases, differences in demographics and health between the covered populations, or even a product of random processes and statistical artefacts. Untangling the factors driving heterogeneity of results is important for helping identify which data sources and results can be given credence and therefore usefully inform health decisions and policy [14].

To help address this issue, comparisons are useful that apply identical methods to two or more independent databases sampling from the same underlying patient population. By keeping the population and methods constant across databases, we can better determine the extent to which the database systems per se produce variability in the results. However, studies of this form are few and far between. Two replication studies using different UK PCDs reported closely corresponding results using different database sources [15-17], but these replications were conducted by research groups instrumental in the creation and maintenance of the comparator PCD and hence lacked independence. In a previous paper [1] we used the Clinical Practice

Research Datalink (CPRD) database [18] to conduct an exact and independent replication of a study originally undertaken in the QResearch database [19] on the impact of statins on survival in patients with coronary heart disease [20]. These databases have no practices in common and use data drawn from different practice electronic record systems (EMIS and VISION respectively). Reassuringly, our results using CPRD were in all main respects identical to those found with QResearch, in particular for the main outcome of overall risk of death associated with statins, which was lower by 55% in CPRD compared to 53% in QResearch.

To further build the evidence base on the validity of studies conducted using UK PCDs, in this paper we report on our second independent replication of a PCD study, comparing results derived from CPRD with results from a previously published study that used another PCD - the Doctors' Independent Network (DIN) [21] - that also does not overlap with CPRD, in either practices or record system. The original study by Shah and colleagues compared all-cause mortality in patients with a new diagnosis of solid cancer receiving beta-blockers with mortality in similar patients receiving alternative antihypertensive medications [22]. This represents a different clinical topic to those addressed by previous replications.

METHODS

The Doctors' Independent Network database (DIN) is an anonymised database drawing data from over 300 general practices using Torex software, covering over 3 million patients since 1989 [21]. There is no overlap between the practices in DIN and those in CPRD. An additional feature that makes DIN appealing for present purposes is that it is built around a quite different philosophy of how the medical record should be structured. CPRD records consultation notes as a sequence of discrete episodes, essentially unconnected, whereas DIN is based around the concept of the Problem Oriented Medical Record (POMR), which treats the medical record as a series of discrete but interconnected problems, with prescriptions linked to diagnoses under problem headings [23].

Data for 1998 for a subset of 142 DIN practices that passed data quality control checks for that year demonstrated very high comparability in age and gender structure to both CPRD and Office for National Statistics mid-year population estimates [21], although DIN practices are somewhat more likely to be located in Southern areas of the UK (Carey, personal correspondence). Prescription records are similar [23, 24] and good agreement has been reported for ischemic heart disease and hay fever prevalence [21] and the recording of thirty common childhood conditions [25].

As in our previous replication, we focused on studies of the effectiveness of medicinal interventions and after assessing the relevant studies that had been conducted in DIN, chose to replicate an investigation into the effects of beta-blocker treatment on cancer survival by Shah and colleagues [22]. This study concerned a quite different patient group and class of drug than our previous replication, and a relatively small treatment effect as opposed to a large one. In addition, the topic under investigation was an incidental drug effect - suggested by earlier in vitro studies [26] - that sparked a great deal of medical community interest and related research activity, still ongoing [27]. The results of this activity have been very mixed and often contradictory, with some studies finding a protective effect for beta-blocker use in relation to mortality from breast cancer [28, 29] and others finding no effect or a modestly increased risk for various cancers, including lung, breast, and prostate [30] and substantially increased risk of developing more advanced colon cancer [31]. Interpretation of this variation in results is not simple as there are many differences between the studies, including the types of beta-blocker involved, which could influence the relationship to mortality [32].

Notwithstanding this complexity, for the purposes of this paper we are primarily concerned with the findings of the particular study by Shah and colleagues, whose DIN-based analysis produced some evidence for an increase in all-cause mortality in cancer patients receiving beta-blockers. The size of the effect across the total sample was small but not insubstantial (an 18% increase in risk of death), with sub-group analyses suggesting that this reflected larger effects mostly confined to patients with pancreatic and prostate cancers and those on beta-blockers without additional blood pressure lowering medications. In their paper, Shah and colleagues acknowledge that they cannot easily explain these results, but conclude that their study does not support the hypothesis that beta- blockers improve survival for common cancers.

Using CPRD, we replicated the methods of Shah and colleagues as closely as possible, given the differences between the two databases. The methodological details provided in the published paper were not sufficient by themselves to allow a close replication to be conducted, and we therefore obtained additional details from the authors. We requested purely factual information about the methods used and did not share any of our analyses or results. All of the methods described below, including the study period, variable specifications and analytical procedures, are exact replications of those used in the original study, unless indicated otherwise.

We selected all practices in CPRD that provided up to standard (UTS) data (UTS is CPRD's designation for data meeting their internal quality standards) for the whole of the period from 1 January 1997 to 31 December 2006. Within these practices and period, we next identified all patients aged 40-85 with a first diagnosis of a solid tumour of the breast, lung, stomach, oesophagus, colon, renal system, prostate or ovary, and with at least two prescriptions of an anti-hypertensive drug (beta-blockers, ACE inhibitors, angiotensin receptor blockers, thiazides, calcium channel blockers, alpha-adrenoceptor blockers) in the year prior to diagnosis. We excluded patients with specific indications (coronary heart disease, heart failure, arrhythmias, stroke) or contra-indications (COPD, diabetes, asthma, renal disease) for antihypertensive medication that may impact on survival, prior to cancer diagnosis. Indications were determined using the Read code lists for the original study as provided to us by Shah and colleagues. We then classified the remaining patients according to exposure in the 1-year period prior to cancer diagnosis into three groups: (i) beta-blockers plus other blood-pressure lowering medications (BPLM); (ii) beta- blockers but no other BPLM; (iii) other BPLM only (controls). All Read codes used in the study are available on the ClinicalCodes repository at https://www.clinicalcodes.org [33].

We extracted data for these patients from 1 year prior to cancer diagnosis, up until the end of 2007, or until the last recorded date for practices that stopped providing data before the end of 2007, giving a maximum possible length of follow-up post-diagnosis of 10 years. We intentionally made no attempt to 'improve' on the analysis conducted by Shah and colleagues as our specific aim was to determine whether the same results and conclusions would emerge from using identical methods on a different underlying dataset.

Analysis

The main outcome was all cause mortality, identified through a record of death in the CPRD. Patients who left their practice during the follow-up period were treated as censored observations in the analysis. Analysis used a Cox proportional hazards model with adjustment for patient age (below 55, 55-65, 66-75, 76 or older), gender, year of diagnosis, smoking status (current, ex-smoker, never smoked, not recorded, as recorded in the year prior to diagnosis), number of medications received in year prior to diagnosis, Regional Health Authority, and practice postcode Index of Multiple Deprivation [34]. The only measure not defined in

the same way as by Shah and colleagues was deprivation, which was at the patient level in DIN but the practice in CPRD (see below).

Again following Shah and colleagues, we conducted an analysis for each cancer site separately, and then combined across sites using a DerSimonian-Laird random-effects meta-analysis. Analyses were undertaken to compare all patients receiving beta- blockers with the controls, and also for patients subdivided into those receiving, and those not receiving, additional BPLM. We further undertook analyses for non-selective beta-blockers only. Missing patient demographic information was dealt with by adding a category for "missing" to the levels of the variable (see Table 1). For all event variables (ie cancer diagnoses, prescriptions, deaths) absence of a relevant code in CPRD was taken to indicate no such event.

To make direct comparisons of the overall and cancer-specific treatment effect estimates (hazard ratios) from CPRD with those reported by Shah and colleagues for DIN, we used a Wald test, computed as the difference between the natural logs of the two hazard ratios divided by the standard error (derived from the logged confidence interval limits), tested as a Z-score [35].

All analyses were performed using R version 3.0.2 [36]. In line with Shah and colleagues, we used an alpha level for statistical significance of 5% throughout.

Sensitivity analyses

We repeated the sensitivity analysis in the original study by excluding patients with less than 1-year survival after cancer diagnosis. The definitions of death and deprivation differed between databases and to assess sensitivity to this we repeated the analyses with the CPRD sample restricted to practices for which linkages to patient-level IMD scores and ONS official death dates were available (58% of practices covering 60% of patients).

We observed notable differences between the cohorts regarding cancer site prevalence rates, area deprivation, year of diagnosis, and patient gender (table 1). Some of these differences, particularly year of diagnosis, are likely related to a considerable increase in the number of practices in the CPRD over the time of the study, compared to DIN (Supp table S1). To examine the sensitivity of our results to these database differences we performed a sensitivity analysis on samples of the CPRD data that matched the make-up of the DIN cohort in key aspects.

To do this, we used an Iterative Proportional Fitting (IPF) [37] algorithm, a method for matching marginal distributions that does not assume independence between the matching variables. The method is described in detail in supplementary File1. The matching variables were cancer site prevalence, year of diagnosis and area deprivation. The algorithm calculated selection probabilities (weights) for each patient in the CPRD data that were used to draw 10,000 weighted bootstrap samples (i.e. samples with replacement). Each sample was analysed and the results combined to obtain overall estimates of effect (the median hazard ratio) and 95% confidence intervals (2.5 and 97.5 percentiles). The IPF algorithm produced an excellent level of agreement on all three matching variables and also corrected the imbalance on gender but not smoking status (Supp table S2). On average, each bootstrap sample consisted of 1,352 patients on beta-blockers and 2,753 on other BPLMs. We also ran an analysis adjusting for the clustering of patients within practices, which Shah and colleagues performed but did not report as it made no difference to results (personal correspondence).

RESULTS

Comparison of patient cohorts

Table 1 compares the patient cohorts from CPRD and DIN on key measures. As expected from the greater number of practices in CPRD, the total sample was much larger (11302 from 582 practices versus 3462 from 171 practices). Patients in the CPRD cohort were more likely to be male (55% v 47%), to be an ex-smoker (44% v 27%), to live in a more deprived area (38% v 25% in the two most deprived quintiles) and to have been more recently diagnosed with cancer (69% v 51% since 2003). CPRD patients were also more likely to have a higher number of recorded medications (56% v 44% on 10 or more medications), though the rate of beta-blocker prescription in CPRD was a little lower (36% versus 41%). The breakdown of types of beta-blockers used was similar in both cohorts, with around three-quarters of patients on atenolol (based on their last prescription before cancer diagnosis): this minimises the risk that the comparison might be affected by differential associations between mortality and beta-blocker type [32].

There was a much higher rate of prostate cancer in the CPRD cohort (Table 2: 33% v 22%) but lower rates of ovarian and renal cancers – though absolute numbers of these were low in both cohorts. Rates for other types of cancer were all similar. The overall mortality rate was also considerably higher in CPRD (Table 2: 50% v 42%), though median lengths of follow-up were similar (29 v 30 months), as were survival rates 1 year post-diagnosis (78% v 74%). The disparity in overall mortality rates was not resolved by matching the CPRD and DIN samples (Table S2: 55% v 42%), nor was it resolved after further matching the samples on smoking status (51% v 42%), and hence cannot be attributed to cohort differences in cancer rates, year of diagnosis, or patient demographic factors. The overall mortality rate was also not reduced in the CPRD sensitivity dataset restricted to patients with ONS deaths and patient level IMD scores (51%).

Survival analysis (Table 3; figure 1)

Overall mortality

There was no difference in adjusted mortality rates between patients in the CPRD receiving beta-blockers (with or without other BPLM) and those on other BPLMs only (overall hazard ratio=1.01, 95%CI 0.91 to 1.13). This compares to a small but statistically significant impact of beta-blockers on mortality in DIN (hazard ratio = 1.18, 95%CI 1.04 to 1.33). However, the Wald test directly comparing these hazard ratios was not statistically significant, although it did approach significance (p=0.063).

For the sub-set of patients on beta-blockers alone, we again found no significant impact on cancer mortality (HR=0.95, 95%CI 0.83 to 1.09), as opposed to a significant effect (HR=1.37, 95%CI 1.16 to 1.61) reported by Shah and colleagues. In this instance the comparison between studies was significant (p<0.001). For the remaining two subsets, of patients on beta-blockers plus other BPLM and patients on non-selective beta-blockers, the two studies returned similar, non-significant, results.

Mortality for individual cancer sites (Table 4)

Using CPRD, mortality rates for patients receiving beta-blockers, compared to those on other BPLMs only, were significantly higher for breast cancer (HR=1.19, 95%CI 1.03 to 1.37) and oesophageal cancer (HR=1.27, 95%CI 1.01 to 1.59) but significantly lower for patients with colon (HR=0.85, 95%CI 0.74 to 0.97) and renal cancer (HR 0.46, 95%CI 0.26 to 0.83), with no significant differences for other cancer sites. Using DIN, Shah and colleagues reported survival to be significantly poorer for patients with pancreas and prostate cancer, with no other differences. Thus for four of the nine cancer sites our CPRD study found a significant

association of mortality with beta-blockers whereas the DIN study did not, and for two other sites this was reversed.

Direct comparison of the cancer site-specific hazard ratios from the two studies using Wald tests found no significant differences except for pancreatic cancer (p=0.023) and prostate cancer (p=0.016). For both cancers CPRD returned hazard ratios close to 1 whereas DIN produced much higher values. There was also significant heterogeneity of treatment effect across cancer sites in CPRD (p=0.004) in contrast to non-significant heterogeneity in DIN (p=0.41).

Results of sensitivity analysis

Sensitivity analysis using the subset of CPRD practices for which ONS mortality and patient-level IMD scores were available produced little change in the overall hazard ratio for death associated with beta-blockers (Table 4: HR=1.09, 95% CI 0.94 to 1.26). Hazard ratios for individual cancer sites likewise did not change greatly, although those for colon and renal cancers ceased to be statistically significant, at least partly due to the reduced sample.

Analysis of the CPRD sample(s) selected to match Shah and colleagues' DIN cohort resulted in an overall hazard ratio and 95% confidence interval identical to our primary analysis, and only small changes in the results for individual cancer sites, although the hazard ratios for oesophageal and renal cancers ceased to be statistically significant (table 4), as did the direct comparison of the CPRD and DIN hazard ratios for pancreatic cancers (p=0.069). Repeating our analyses adjusting for clustering of patients within practices and excluding patients who survived for less than a year made no substantive difference to any of the results (Table S3).

DISCUSSION

We conducted a fully independent, external, replication of a study based on one PCD using data from an alternative database. Our replication used the Clinical Practice Research Datalink (CPRD), a larger dataset than the Doctors' Independent Network (DIN), hence our total patient sample was more than three times the size of the original study. As far as we were able, we sampled from the same patient population and used identical methods to the original study, to minimise any sources of variation other than the database itself.

Using CPRD we found no evidence for an association between beta-blocker use and cancer mortality, either in the full CPRD cohort or for patients on beta-blockers only - where the strongest effect was observed by Shah and colleagues using DIN. Results for individual cancer types also differed considerably, indicating an entirely different set of statistically significant cancer sites. However, most study differences disappeared under direct comparison of the hazard ratio estimates, with only the treatment effects for patients with pancreatic and prostate cancers and for those on beta-blockers alone remaining significantly different, for whom beta-blocker use was associated with mortality in DIN but not CPRD. Thus with these exceptions, all treatment effect estimates from the two studies agreed within the range of random variation. These results were unchanged in all essentials under sensitivity analyses using CPRD subsamples with linked ONS mortality and patient-level deprivation measures, and matched to DIN on cancer prevalence rates and other sample characteristics.

It is informative to compare both of these studies to a series of investigations by a group centred at Queen's University, Belfast, who also used CPRD to investigate the effects of beta-blocker usage on mortality from

breast [27], colon [38] and prostate cancer [39], using a methodology that differed in a number of respects. No significant associations were found for any of these cancer sites, in contrast to both ourselves (breast, colon) and to Shah and colleagues (prostate). However, the confidence intervals reported by all three teams overlapped considerably - with the exception of prostate cancer from Shah and colleagues (table S4) – indicating that all treatment effect estimates were equivalent within statistical limits of accuracy.

Unresolved differences between CPRD and DIN after direct comparison were few and mainly low-level, but it is worthwhile to consider why these should remain. Results did show some sensitivity to database differences in patient demographics, though not enough to explain all the discrepant results. The very different clinical computing systems used may have affected aspects of recorded care – possibly more than any practice or sample characteristic [40] - but the small number of study differences suggests that any overall impact was minimal, though an influence on specific data items and results is plausible. The much lower mortality rate in DIN suggests that death may have been defined differently, or recorded less reliably, but for this to explain the difference in effect estimates, the act of recording mortality in DIN would have to be associated with prescription of beta-blockers but not prescription of other BPLMs, and only for certain cancers but not others, which seems unlikely.

The unresolved differences might simply be statistical artefacts. Unmeasured confounding factors could vary in distribution between the datasets. Also, all the discrepant results concerned largely exploratory sub-group analyses done within a framework of multiple significance testing and arguably an alpha-level higher than 5% would be more appropriate: at alpha=1%, the only unresolved difference is for patients on beta-blockers alone. The QResearch replication study [20] likewise identified a number of within- and between-study discrepancies in subgroup analyses, relating to different statin compounds, and in all three of the Belfast group's studies, despite no overall associations sub-group analyses found significant relationships between cancer survival and from one to three specific beta-blocker compounds, though not always the same compound and not always in the same direction [27, 38, 39]. The large scale of many EHRs may encourage researchers to undertake multiple subgroup analyses without any firm hypotheses, and may also foster the idea that size alone offers some protection against incorrect inference, yet the rate at which inconsistent results occur in EHR-based studies strongly suggests that issues of multiple testing, "fishing" for results, and spurious significance apply as much to these data sources as they do to much smaller datasets; possibly even more so given the potential for bias from residual and uncontrolled confounding.

The results of our study therefore present a somewhat complex picture: examined separately and purely in terms of statistical significance, the CPRD- and DIN-based studies provided rather different pictures of the risks of beta-blockers overall and in relation to different cancer types. The survival disadvantage observed by Shah and colleagues was not insubstantial: an increased point risk of death of 18%, increasing to 37% for patients on beta-blockers only. These results were not present in our replication study, including across a variety of sensitivity analyses. Yet when directly compared, with the main exception of the beta-blocker only subgroup estimates of treatment effect from the two studies did not differ statistically. Drawing a satisfactory conclusion from these findings is not easy. Focusing on the direct statistical comparisons of the study effect estimates, this study taken in combination with our previous replication study and other non-independent replications, suggests that the application of identical analytical methods to different UK PCDs yields treatment effect estimates that are usually comparable within statistical limits of accuracy. Nevertheless, taken separately our study and that of Shah and colleagues point to very different conclusions about the safety of beta-blockers in this patient population, indicating that single studies, even when

demonstrating notable effects based on well-substantiated databases, do not guarantee generalisable results.

Limitations

Differences between the DIN and CPRD databases meant that while we were able to exactly replicate the great majority of the components of the original study, there were a few exceptions. The datasets may have differed in their definitions of all-cause mortality, as each use their own bespoke algorithm. For area deprivation, Shah and colleagues used 2004 IMD scores in national quintiles based on each patient's postcode. Equivalent scores were only available to us for a subset of CPRD, so instead we used 2004 practice-postcode IMD scores, obtained for all practices from the CPRD organisation as a linked dataset. We tested for the impact of these factors by running a sensitivity analysis using the subset of CPRD patients for which linked ONS data on the date of death and residential IMD 2004 scores were available. In all other respects, this study replicated the original with respect to the population and variable definitions and methods of analysis.

The overall raw mortality rate in our CPRD cohort was substantially higher than in the DIN cohort and a much higher proportion of CPRD cancers were of the prostate. Patients in the CPRD cohort were also likely to have been diagnosed more recently, to live in areas of higher deprivation and to be male. However, analysis of subsets of the CPRD cohort matched to DIN did not account for the difference in overall mortality rates, nor did it substantially alter our findings. Neither the complete details of how Shah and colleagues defined mortality nor of the CPRD mortality algorithm were available to us, thus our ability to uncover the reasons for these different mortality rates was limited. However, sensitivity analysis using ONS official mortality data suggested that the CPRD mortality rates at least are robust.

We intentionally did not try to improve on the analysis methods used by Shah and colleagues, even though these have received some criticism [32, 38], since for our purposes it was important to keep the analysis methods constant. Criticisms include: not linking to cancer registries; lack of control for stage of disease or treatment; not differentiating beta-blocker use prior- and post- cancer diagnosis; and use of patients on other antihypertensives as the comparator. Most of these criticisms were in fact discussed by Shah and colleagues in their paper and justified there as part of the methods. Importantly, the Belfast group's CPRD-based studies took account of most of these issues yet yielded effect estimates very similar to our own.

Conclusion

This replication of one UK PCD-based study in a second completely independent PCD using the same methods and sampling the same population has revealed a complex pattern of similarities and differences in both the make-up of the patient cohorts and in the findings from analysis. When directly compared, with the exception of certain subgroup results, estimates of treatment effect did not differ statistically and in this sense this study adds to previous replication work in finding that when analysed to a common protocol, different UK PCDs produce treatment effect estimates that generally agree within statistical tolerance. Nevertheless, considered separately, this study and the original DIN-based investigation point to very different conclusions regarding the safety of beta-blockers for solid cancer patients. Hence our results also show that single studies based on even these internally well-validated databases may not guarantee generalisable results. Therefore great care must be taken in drawing any firm conclusions, particularly where subgroup results are concerned. In all cases, confirmatory studies using at least one other independent data source are strongly recommended.

REFERENCES

1. Reeves D, Springate DA, Ashcroft DM, et al. Can analyses of electronic patient records be independently and externally validated? The effect of statins on the mortality of patients with Ischaemic heart disease: a cohort study with nested case-control analysis. *BMJ Open* 2014;4:e004952

2. Herrett E, Thomas SL, Schoonen WM, et al. Validation and validity of diagnoses in the General Practice Research Database: a systematic review. *Br J Clin Pharmacol* 2010;69:4-14.

3. Khan NF, Harrison SE, Rose PW. Validity of diagnostic coding within the General Practice Research database: a systematic review. *BJGP* 2010;e128-e136.

4. Jordan K, Porcheret M, Croft P. Quality of morbidity coding in general practice computerized medical records: a systematic review. *Family Practice* 2004;21:396-412.

5. Marston L, Carpenter JR, Walters KR, et al. Issues in multiple imputation of missing data for large general practice clinical databases. *Pharmacoepidemiol Drug Saf* 2010;19:618-626.

6. Delaney JA, Daskalopoulou SS, Brophy JM, et al. Lifestyle variables and the risk of myocardial infarction in the General Practice Research Database. *BMC Cardiovascular Disorders* 2007;7:38.

7. Collins SC, Altman DG. An independent and external validation of QRISK2 cardiovascular disease risk score: a prospective open cohort study. BMJ 2010 340:c2442.

8. Madigan D, Ryan PB, Schuemie M, et al. Evaluating the Impact of Database Heterogeneity on Observational Study Results. *Am J Epidemiol* 2013;178:645-651.

9. Wijnans L, Lecomte C, de Vries C, et al. The incidence of narcolepsy in Europe: Before, during, and after the influenza A(H1N1)pdm09 pandemic and vaccination campaigns. *Vaccine* 2013;31:1246-1254.

10. Dormuth CR, Hemmelgarn BR, Paterson JM, et al. Use of high potency statins and rates of admission for acute kidney injury: multicenter, retrospective observational analysis of administrative databases. *BMJ* 2013;346:f880

11. Filion KB, Chateau D, Targownik LE, et al. Proton pump inhibitors and the risk of hospitalisation for community-acquired pneumonia: replicated cohort studies with meta-analysis. *Gut* 2014;63:552-8.

12. Valkhoff VE, van Soest EM, Mazzaglia G, et al. Adherence to gastroprotection during cyclooxygenase 2 inhibitor treatment and the risk of upper gastrointestinal tract events: a population-based study. *Arthritis Rheum* 2012;64:2792-2802.

13. Andrews N, Stoweb J, Miller E, et al. A collaborative approach to investigating the risk of thrombocytopenic purpura after measles–mumps–rubella vaccination in England and Denmark. *Vaccine* 2012;30:3042-3046.

14. Bell C, Chakravarty A, Gruber S, et al. Characteristics of study design and elements that may contribute to the success of electronic safety monitoring systems. *Pharmacoepidemiol Drug Saf* 2014;DOI:10.1002/pds.3712

15. Bremner SA, Carey IM, DeWilde S, et al. Early-life exposure to antibacterials and the subsequent development of hayfever in childhood in the UK: case-control studies using the General Practice Research Database and the Doctors' Independent Network. *Clin Exp Allergy* 2003;33:1518-25.

16. Bremner SA, Carey IM, DeWilde S, et al. Vaccinations, infections and antibacterials in the first grass pollen season of life and risk of later hayfever. *Clin Exp Allergy* 2007;37:512-517.

17. Vinogradova Y, Coupland C, Hippisley-Cox J. Exposure to bisphosphonates and risk of gastrointestinal cancers: series of nested case-control studies with QResearch and CPRD data. *BMJ* 2013;346:f114 doi: 10.1136/bmj.f114.

18. Clinical Practice Research Datalink (CPRD). Available: http://www.cprd.com/home. Date accessed 20 March 2014.

19. QResearch. Available: http://www.qresearch.org. Date accessed 20 March 2014

20. Hippisley-Cox J, Coupland C. Effect of statins on the mortality of patients with ischaemic heart disease: population based cohort study with nested case-control analysis. *Heart* 2006;92:752-758.

21. Carey IM, Cook DG, De Wilde S, et al. Developing a large electronic primary care database (Doctors' Independent Network) for Research. *Int J Med Inform* 2004;73:443-453.

22. Shah SM, Carey IM, Owen CG, et al. Does b-adrenoceptor blocker therapy improve cancer survival? Findings from a population-based retrospective cohort study. *Br J Clin Pharmacol* 2011;72:157–161.

23. Carey IM, Cook DG, De Wilde S, et al. Implications of the problem orientated medical record (POMR) for research using electronic GP databases: a comparison of the Doctors Independent Network Database (DIN) and the General Practice Research Database (GPRD). *BMC Fam Pract* 2003;4: 14.

24. De Wilde S, Carey IM, Bremner SA, et al. Evolution of statin prescribing 1994-2001: a case of agism but not of sexism? *Heart* 2003;89:417-421.

25. De Wilde S, Carey IM, Bremner S. A comparison of the recording of 30 common childhood conditions in the Doctor's Independent Network and General Practice Research Databases. *Health Stat Q* 2004;22:21–31.

26. Chakroborty D, Sarkar C, Basu B, et al. Catecholamines regulate tumor angiogenesis. *Cancer Res* 2009;69:3727–30.

27. Cardwell CR, Coleman HG, Murray LJ, et al. Beta-blocker usage and breast cancer survival: a nested case-control study within a UK Clinical Practice Research Datalink cohort. *Int J Epidemiol* 2013;42:1852-61 doi: 10.1093/ije/dyt196.

28. Ganz PA, Habel LA, Weltzien EK, et al. Examining the influence of beta blockers and ACE inhibitors on the risk for breast cancer recurrence: results from the LACE cohort. *Breast Cancer Res Treat* 2011;129:549–556.

29. Melhem-Bertrandt A, Chavez-Macgregor M, Lei X, et al. Beta-blocker use is associated with improved relapse-free survival in patients with triple-negative breast cancer. *J Clin Oncol* 2011;29:2645–2652.

30. Friedman GD, Udaltsova N, Habel LA. Norepinephrine antagonists and cancer risk. *Int J Cancer* 2011;128:737–738.

31. Jansen L, Below J, Chang-Claude J, et al. Beta blocker use and colorectal cancer risk: population-based case-control study. *Cancer* 2012;118:3911-3919.

32. Barron TI, Sharp L, Visvanathan K. Beta-adrenergic blocking drugs in breast cancer: A perspective review. *Ther Adv Med Oncol* 2012;4:113–125 doi: 10.1177/1758834012439738

33. Springate DA, Kontopantelis E, Ashcroft DM, et al. ClinicalCodes: An Online Clinical Codes Repository to Improve the Validity and Reproducibility of Research Using Electronic Medical Records. *PLoS ONE* 2014;9:e99825 doi:10.1371/journal.pone.0099825

34. Available: http://data.gov.uk/dataset/imd_2004. Date accessed 20 October 2014.

35. Tierney JF, Stewart LA, Ghersi D, et al. Practical methods for incorporating summary time-to-event data into meta-analysis. *Trials* 2007;8:16

36. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing 2013. Available: http://R-project.org. Date accessed 20 November 2014

37. Speed TP. Iterative Proportional Fitting. *Encyclopedia of Biostatistics* 2005;4:doi: 10.1002/0470011815.b2a10027

38. Hicks BM, Murray LJ, Powe DG, et al. β-Blocker usage and colorectal cancer mortality: a nested case-control study in the UK Clinical Practice Research Datalink cohort. *Ann Oncol* 2013;24:3100-6 doi: 10.1093/annonc/mdt381.

39. Cardwell CR, Coleman HG, Murray LJ, et al. Beta-blocker usage and prostate cancer survival: a nested case-control study in the UK Clinical Practice Research Datalink cohort. *Cancer Epidemiol* 2014;38:279-85 doi: 10.1016/j.canep.2014.03.011.

40. Kontopantelis E, Buchan I, Reeves D, et al. Relationship between quality of care and choice of clinical computing system: retrospective analysis of family practice performance under the UK's quality and outcomes framework. *BMJ Open* 2013;3:e003190 doi:10.1136/bmjopen-2013-003190

Acknowledgements
This study is based on data from the Clinical Practice Research Datalink (CPRD) obtained under licence from the UK Medicines and Healthcare products Regulatory Agency. However, the interpretation and conclusions contained in this paper are those of the authors alone. We are grateful to Dr Iain M. Carey from the Division of Population Health Sciences and Education, St George's University of London, for his comprehensive responses to our queries about the original DIN study by himself, Shah and colleagues and about DIN itself.

Contributors
DR, EK, and RR developed the original idea for the study and DS, DA and TD contributed to the study design. DR supervised all aspects of the study's execution. DS helped to plan the analysis and undertook the primary analysis. DR and DS wrote the first draft of the paper. All authors critically reviewed the paper and approved the submitted version. DS had full access to all the data in the study and he takes responsibility for the integrity of the data and the accuracy of the data analysis.

Funding
This work was funded by the National Institute for Health Research School for Primary Care Research (project number 142). The funders had no role in the conducting of the study or in the decision to submit the paper for publication. The views expressed are those of the authors and not necessarily those of the NHS, the National Institute for Health Research or the Department of Health.

Competing interests
All authors have completed the Unified Competing Interest form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare that DR was partly funded and DS fully funded by an NIHR School for Primary Care Research grant to undertake this study and EK was partly supported by a NIHR School for Primary Care Research fellowship in primary health care but that all other authors received no support from any additional organisation for the submitted work and had no relationships with any additional organisation that might have an interest in the submitted work in the previous 3 years. For all authors, their spouses, partners, or children have no financial relationships that may be relevant to the submitted work; and no non-financial interests that may be relevant to the submitted work.

Ethics approval
The study protocol was approved by the independent scientific advisory committee (ISAC) for CPRD research (reference number: 12_149R).

Data sharing
Clinical Practice Research Datalink data cannot be shared due to licencing restrictions. All the clinical code-lists used in the analysis of CPRD in this study are available at https://www.clinicalcodes.org/.The full R code used for the analysis of CPRD is available from the authors.

Table 1. Comparison of CPRD and DIN patient cohorts (%, (n))

Variable		CPRD (n=11302)	DIN (n=3462)
Age at diagnosis (years)	18-55	5.3% (602)	NA
	56-65	23.3% (2631)	NA
	66-75	37.4% (4228)	NA
	75 and above	33.9% (3841)	NA
Gender	Male	55.3% (6247)	47.4% (1641)
	Female	44.7% (5055)	52.6% (1821)
Smoking	Current	14.7% (1665)	19.1% (661)
	ex-smoker	43.9% (4962)	27.2% (941)
	Never smoked	34.2% (3864)	51.8% (1792)
	Missing	7.2% (811)	1.9% (68)
Deprivation (IMD 2004 Quintiles)[a]	1 (most deprived)	17.8% (2012)	9.6% (333)
	2	19.9% (2253)	14.9% (517)
	3	21% (2375)	19.2% (664)
	4	23.1% (2613)	22.0% (760)
	5 (least deprived)	18.1% (2049)	26.4% (915)
	Missing	0% (0)	7.9% (273)
Year of diagnosis	1997-8	5.8% (658)	12.1% (420)
	1999-00	8.8% (996)	15.8% (546)
	2001-2	16.4% (1856)	20.6% (714)
	2003-4	29.2% (3303)	25.1% (870)
	2005-6	39.7% (4489)	26.3% (912)
N of medications	0-4	14.9% (1681)	16.9% (586)
	5-9	29.4% (3319)	38.1% (1318)
	10-14	28.2% (3188)	24.4% (845)
	15-19	8.5% (956)	9.6% (332)
	20 and above	19.1% (2158)	7.8% (269)
	Missing	0% (0)	3.2 (112)
Prescribed b-blocker	No	64.3% (7272)	59.4% (2057)
	Yes	35.7% (4030)	40.6% (1405)
Type of beta-blocker	Atenolol	73.0% (2943)	75.2% (1057)
	Propranolol	11.0% (443)	12.8% (180)
	Other beta-blocker	16.0% (644)	12.0% (168)

NA = Not Available

[a]Based on patient postcode for DIN and practice postcode for CPRD

Table 2. Comparison of CPRD and DIN patient cohorts by exposure to blood pressure lowering medication (BPLM) in the year prior to cancer diagnosis

	All patients				Other BPLMs but no beta-blockers				Beta-blockers plus other BPLMs				Beta-blockers only (beta-blockers but no other BPLMs)			
	CPRD		DIN		CPRD		DIN		CPRD		DIN		CPRD		DIN	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
All patients	11302	100	3462	100	7272	64.3	2056	59.4	2832	25.10	864	24.4	1198	10.6	542	15.7
Deaths	5754	50.1	1441	41.6	3748	51.6	846	41.2	1459	51.5	350	40.5	547	45.7	245	45.2
Alive at 1-year follow-up	8763	77.5	2576	74.4	5607	77.1	1541	75.0	2194	77.5	630	72.9	962	80.3	405	74.7
On non-selective beta-blocker	685	6.06	239	7.5	NA	NA	NA	NA	359	12.7	71	8.2	326	27.2	167	30.8
Cancer sites																
Breast	2943	26.0	984	28.4	1746	24.0	554	26.9	794	28.0	240	27.8	403	33.6	194	35.8
Colon	1799	15.9	619	17.9	1104	15.2	354	17.2	476	16.8	162	18.7	219	18.3	103	18.9
Lung	1326	11.7	436	12.6	913	12.6	277	13.5	307	10.9	105	12.1	106	8.8	54	9.9
Oesophagus	434	3.8	159	4.6	257	3.5	95	4.6	116	4.1	44	5.1	61	5.1	20	3.7
Ovarian	203	1.8	148	4.3	124	1.7	76	3.7	45	1.6	43	5.0	34	2.8	29	5.4
Pancreas	376	3.3	140	4.0	222	3.1	83	4.0	111	4.0	34	3.9	43	3.6	23	4.2
Prostate	3748	33.2	759	21.9	2604	35.8	500	24.3	856	30.2	182	21.0	288	24.0	77	14.1
Renal	141	1.2	124	3.6	81	1.1	69	3.4	48	1.7	33	3.8	12	1.0	22	4.0
Stomach	332	2.9c	93	2.7	221	3.0	52	2.5	79	2.8	21	2.4	32	2.7	20	3.7

NA = Not Applicable

Table 3. Comparison of patients using beta-blockers versus patients using other BPLMs only from the CPRD and DIN studies: pooled hazard ratios (95% CI)) from meta-analyses of cancer site-specific results

Comparison (versus controls)	CPRD	DIN	*p*-value†
All patients using beta-blockers	1.01 (0.91, 1.13)	1.18 (1.04 to 1.33)**	0.06
Patients using beta-blockers only	0.94 (0.82, 1.07)	1.37 (1.16, 1.61)***	<0.001***
Beta-blockers and other BPLM	1.06 (0.95, 1.19)	1.11 (0.91, 1.34)	0.69
Non-selective beta-blockers only	0.96 (0.8, 1.15)	1.21 (0.94, 1.55)	0.14

†Wald tests of CPRD vs DIN hazard ratios

**p<0.01; ***p<0.001

Table 4. Cancer site-specific and pooled hazard ratios (95% CI)) from CPRD and DIN studies for all patients using beta-blockers versus patients using other BPLMs only

Cancer site	CPRD full cohort	CPRD sensitivity analysis 1: ONS mortality and & deprivation	CPRD sensitivity analysis 2: matched CPRD and DIN samples	DIN primary analysis	*p*-value †
Sample size (on beta-blockers; controls)	4030; 7272	2528; 4514	4105[a]; 7197[a]	1406; 2056	-
Cancer sites					
Breast	1.19 (1.03, 1.37)*	1.28 (1.07, 1.32)***	1.24 (1.09, 1.43)**	1.09 (0.80, 1.49)	0.62 (0.46)
Colon	0.85 (0.74, 0.97)*	0.9 (0.76, 1.07)	0.87 (0.76, 1.0)*	1.00 (0.77, 1.30)	0.28 (0.36)
Lung	1.04 (0.91, 1.19)	0.97 (0.82, 1.15)	1.03 (0.88, 1.21)	1.12 (0.89, 1.41)	0.59 (0.56)
Oesophagus	1.27 (1.01, 1.59)*	1.65 (1.22, 2.24)***	1.19 (0.93, 1.52)	1.05 (0.69, 1.60)	0.44 (0.61)
Ovarian	1.05 (0.74, 1.5)	1.5 (0.89, 2.52)	1.02 (0.77, 1.33)	1.14 (0.63, 2.06)	0.82 (0.74)
Pancreas	0.94 (0.74, 1.21)	0.82 (0.6, 1.11)	1.04 (0.75, 1.43)	1.88 (1.09, 3.25)*	0.023* (0.069)
Prostate	1.03 (0.92, 1.15)	1.02 (0.88, 1.17)	0.94 (0.82, 1.08)	1.54 (1.13, 2.09)**	0.016* (0.004)**
Renal	0.46 (0.26, 0.83)**	0.6 (0.28, 1.27)	0.67 (0.41, 1.03)	1.14 (0.52, 2.52)	0.069 (0.25)
Stomach	1.03 (0.78, 1.36)	1.34 (0.93, 1.93)	1.01 (0.68, 1.48)	1.44 (0.76, 2.74)	0.35 (0.35)
All patients using beta-blockers	1.01 (0.91, 1.13)	1.09 (0.94, 1.26)	1.01 (0.91, 1.13)	1.18 (1.04, 1.33)**	0.063 (0.063)

[a]Median across bootstrap samples

†Wald tests of CPRD full cohort v DIN (CPRD matched sample v DIN)

* $p<0.05$; ** $p<0.01$; *** $p<0.001$

Figure 1. Hazard ratios of survival for patients prescribed beta- blocker therapy compared to patients prescribed other BPLMs

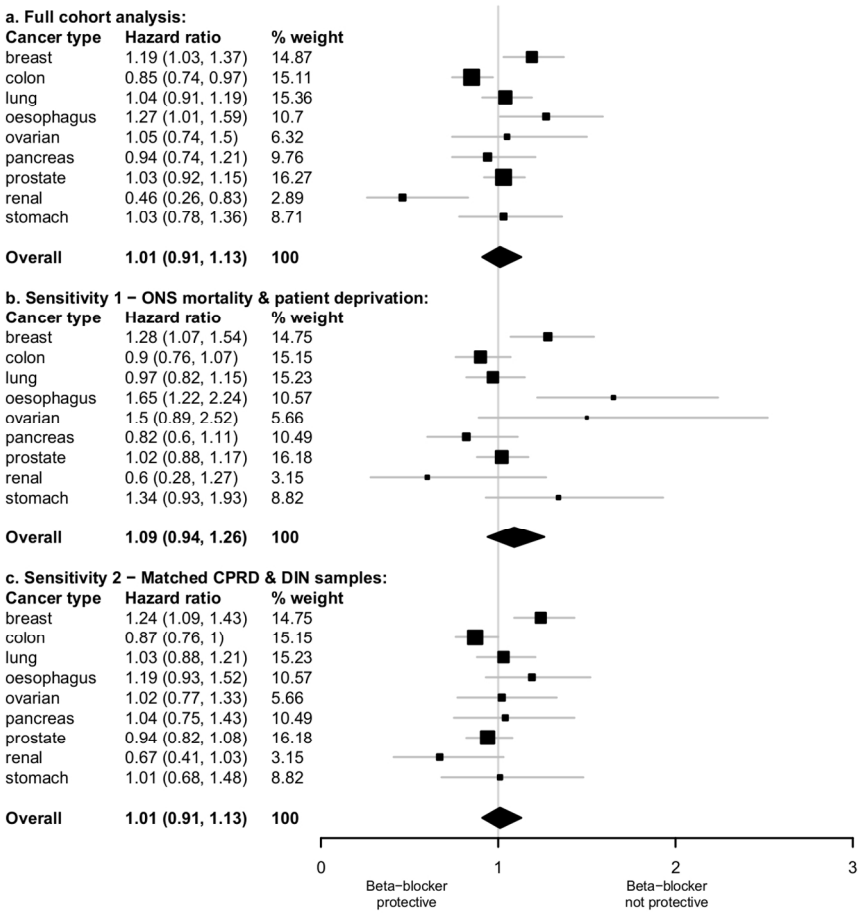

Figure 1. Hazard ratios of mortality for patients prescribed beta- blocker therapy compared to patients prescribed other BPLMs
127x134mm (300 x 300 DPI)

Table S1: Numbers of practices and newly diagnosed patients in each cohort (practices with one or more eligible patients only)

Year	CPRD		DIN	
	N of practices[a]	N of patients	N of practices[a]	N of patients
1997	155	303	88	192
1998	172	355	100	228
1999	200	398	99	253
2000	254	598	119	293
2001	303	808	119	360
2002	350	1048	124	354
2003	425	1427	140	426
2004	495	1876	148	444
2005	510	2090	143	474
2006	526	2399	140	438

[a]The total number of unique practices contributing data was 582 for CPRD and 171 for DIN

Table S2: Comparison of CPRD and DIN patient cohorts, matched samples (%, (n))

Variable		**CPRD matched samples[a] (n=11,302)**	**DIN (n=3462)**
Cancer site	Breast	28.7% (3244)	28.4% (984)
	Colon	18.1% (2042)	17.9% (619)
	Lung	12.7% (1437)	12.6% (436)
	Oesophagus	4.5 (514)	4.6% (159)
	Ovarian	3.8 (427)	4.3% (148)
	Pancreas	4.0 (455)	4.0% (140)
	Prostate	22.1 (2501)	21.9% (759)
	Renal	3.3 (372)	3.6% (124)
	Stomach	2.7 (307)	2.7% (93)
Year of diagnosis	1997-8	11.7% (1321)	12.1% (420)
	1999-00	15.6% (1765)	15.8% (546)
	2001-2	20.7% (2340)	20.6% (714)
	2003-4	25.4% (2868)	25.1% (870)
	2005-6	26.6% (3006)	26.3% (912)
Deprivation (IMD 2004 Quintiles)[b]	1 (most deprived)	10.4% (1175)	10.4% (333)
	2	16.2% (1831)	16.2% (517)
	3	20.8% (2340)	20.8% (664)
	4	23.9% (2871)	23.9% (760)
	5 (least deprived)	28.7% (3006)	28.7% (915)
Gender	Male	47.7% (5393)	47.4% (1641)
	Female	52.3% (5909)	52.6% (1821)
Smoking	Current	15.1% (1711)	19% (661)
	Ex-smoker	45.8% (5173)	27% (941)
	Never smoked	29.6% (3343)	52% (1792)
	Not recorded	9.5% (1075)	2% (68)
Prescribed beta-blocker	No	63.7% (7197)	59.4% (2057)
	Yes	36.3% (4105)	40.6% (1405)
Died	No	44.9% (5082)	58.4% (2021)
	Yes	55.1% (6220)	41.6% (1441)

[a]Counts and %'s for CPRD are medians across all bootstrapped samples

[b]Based on patient postcode for DIN, and practice postcode for CPRD

Table S3: Cancer site-specific and pooled hazard ratios (95% CI)) from CPRD for all patients using beta-blockers versus controls: additional sensitivity analyses

	CPRD sensitivity analysis 3: Clustering of patients by practice	CPRD sensitivity analysis 4: excluding patients who survived for < 1 year
Sample size (on beta-blockers; controls)	4030; 7272	3156; 5607
Cancer sites		
breast	1.19 (1.03, 1.36)*	1.22 (1.05, 1.43)*
colon	0.85 (0.73, 0.98)*	0.91 (0.77, 1.08)
lung	1.04 (0.89, 1.21)	0.97 (0.75, 1.25)
oesophagus	1.27 (0.99, 1.62)	0.99 (0.64, 1.52)
ovarian	1.05 (0.73, 1.51)	1.41 (0.87, 2.29)
pancreas	0.94 (0.71, 1.25)	1.81 (0.84, 3.92)
prostate	1.03 (0.92, 1.15)	1.02 (0.9, 1.16)
renal	0.46 (0.25, 0.85)*	0.63 (0.29, 1.34)
stomach	1.03 (0.77, 1.37)	1.17 (0.67, 2.06)
All patients using beta-blockers	1.01 (0.91, 1.13)	1.05 (0.94, 1.18)

* $p<0.05$

Table S4: Summary of results for breast, colon and prostate cancer for the current study, Shah et al and the Belfast Group (BG)

Cancer type	Current (CPRD): cohort analysis, all-cause mortality (Hazard Ratio)	Shah et al (DIN): cohort analysis, all-cause mortality (Hazard Ratio)	BG[1] (CPRD): cohort analysis, cancer specific mortality (Hazard Ratio)	BG[1] (CPRD): case-control analysis, all-cause mortality (Odds Ratio)
Breast	1.19 (1.03, 1.37)*	1.09 (0.80, 1.49)	0.93 (0.83, 1.06)	1.03 (0.92, 1.16)
Colon	0.85 (0.74, 0.97)*	1.00 (0.77, 1.30)	0.90 (0.79, 1.01)	0.88 (0.77, 1.0)
Prostate	1.03 (0.92, 1.15)	1.54 (1.13, 2.09)**	0.98 (0.86, 1.12)	1.05 (0.94, 1.17)

[1]The BG reported hazard ratios from cohort analysis for cancer-specific death only, but odds-ratios from case-control analysis for all-cause mortality. We therefore include both sets of results in the table.

*$p<0.05$; **$p<0.01$

Iterative proportional fitting for matching on population characteristics between CPRD and DIN-LINK

David A Springate

2015-03-09

Introduction

Iterative Proportional fitting (IPF) is a mathematical scaling method for combining information from two (or more) datasets. It can be used to ensure that a table of data is adjusted so that its marginal (row and column) totals agree with constraining marginal totals obtained from an alternative source. IPF acts as a weighting system whereby the original table values are gradually adjusted through repeated (iterative) calculations to fit these marginal constraints. The resultant table of data is a joint probability distribution of maximum likelihood estimates obtained when the probabilities are convergent within an acceptable (pre-defined) level of tolerance (Norman 1999; Speed 2005). IPF has been proven to converge to the unique maximum likelihood estimates for table cell values, given the user-provided constraints (Fienberg 1970). An advantage of this method is that it allows matching on all variables by adjusting the marginal probabilities whilst preserving the interaction structure of the original table. The algorithm also converges very quickly, typically in less than five iterations, even with multiple variables and cross-tabulated cells.

In this paper, we used IPF to match distributions of patients in CPRD to their corresponding distributions in DIN-LINK according to three matching variables:

1. Cancer site prevalence (9 levels)
2. Year of diagnosis (10 levels)
3. Area deprivation (5 levels)

This is because we observed quite different proportions of patients in the two databases for these variables (See tables 1 and 2 in the main paper) and we wanted to investigate whether the results we obtained were influenced by these different proportions. IPF allowed us to analyse repeated CPRD sub-cohorts matching the marginal proportions of the above three variables with those in the DIN cohort.

The method

We used a two-stage approach where marginal probabilites were first generated by IPF and then applied to a weighted bootstrapping algorithm to generate samples with replacement from the CPRD dataset closely matching the DIN marginal distributions of the above variables.

Stage 1: The IPF algorithm

1. A tolerance, *tol* is set
2. A cross-tabluation of the levels of the three matching variables is constructed (9 x 10 x 5 = 450 cells)
3. A weight is assigned to each of the 450 cells by multiplying together the marginal probabilities from the CPRD cohort of the three matching variables in each cell
4. The marginal CPRD probabilities are recalculated by cross-tabulating the weights
5. New weights are calculated by multiplying the existing weights by the original marginal probabilities in DIN divided by the recalculated CPRD marginal probabilities, for each of the matching variables

6. The maximum absolute difference, D, between the original and new weights is found
7. The original weights are replaced by the new weights
8. If $D > tol$, repeat steps 4-6

Stage 2: The weighted bootstrap

Each patient in the original cohort is assigned a weight, generated from the IPF algorithm, according to their particular combination of matching variables. Then, for each of 10,000 iterations:

1. A sample with replacement of the full original cohort is taken, weighted by the assigned IPF weights generated above
2. Survival analysis is run on the sample to generate hazard ratios in the same way as for the main cohort

After 10,000 replicates, the results are pooled and the median and 2.5 and 97.5 percentiles of the hazard ratios and of the proportions of unmatched variables such as smoking and gender are taken for comparison with the original analysis.

Example

Here follows an illustrative example of the IPF algorithm in R code. We want to match the marginal proportions of two three-level variables, A and B from dataset $d1$ to dataset $d2$:

```
tol = 0.001                              # tolerance for weight difference
d1_A <- c(a1 = 38, a2 = 52, a3 = 55)     # Marginal counts for A in d1
d1_B <- c(b1 = 55, b2 = 42, b3 = 48)     # Marginal counts for B in d1
d2_A <- c(a1 = 80, a2 = 134, a3 = 46)    # Marginal counts for A in d2
d2_B <- c(b1 = 60, b2 = 68, b3 = 132)    # Marginal counts for B in d2

## convert to marginal probabilities:
d1_pA <- d1_A / sum(d1_A)
d1_pB <- d1_B / sum(d1_B)
d2_pA <- d2_A / sum(d2_A)
d2_pB <- d2_B / sum(d2_B)

## Cross-tabulate and assign weights
prob_table <- expand.grid(A = names(d1_pA),
                          B = names(d1_pB))
prob_table$weights <- sapply(1:nrow(prob_table),
                             function(x){
                               d1_pA[[prob_table$A[x]]] *  d1_pB[[prob_table$B[x]]]
                             })

## Add marginal probabilities for target d2 to prob_table:
prob_table$d2_pA <- sapply(prob_table$A, function(x) d2_pA[x])
prob_table$d2_pB <- sapply(prob_table$B, function(x) d2_pB[x])
```

```
## display the table so far:
prob_table
```

```
##    A  B    weights     d2_pA     d2_pB
## 1 a1 b1 0.09940547 0.3076923 0.2307692
## 2 a2 b1 0.13602854 0.5153846 0.2307692
## 3 a3 b1 0.14387634 0.1769231 0.2307692
## 4 a1 b2 0.07590963 0.3076923 0.2615385
## 5 a2 b2 0.10387634 0.5153846 0.2615385
## 6 a3 b2 0.10986920 0.1769231 0.2615385
## 7 a1 b3 0.08675386 0.3076923 0.5076923
## 8 a2 b3 0.11871581 0.5153846 0.5076923
## 9 a3 b3 0.12556480 0.1769231 0.5076923
```

```
## define a weighting function for IPF.
## Note that d1_pA and d1_pB are assigned inside the loop below
reweight <- function(prob_table){
  weights <- prob_table$weights
  for(v in c("A", "B")){
    weights <- weights * (prob_table[[paste0("d2_p", v)]] /
                          prob_table[[paste0("d1_p", v)]])
  }
  weights
}

## Iterate until the weights do not change within the tolerance
iter <- 1 # counter
repeat {
  ## assign marginal probabilities for d1
  prob_table$d1_pA <- sapply(prob_table$A,
                             function(x) xtabs(weights ~ A, data = prob_table)[x])
  prob_table$d1_pB <- sapply(prob_table$B,
                             function(x) xtabs(weights ~ B, data = prob_table)[x])
  ## Recalculate weights
  weights <- reweight(prob_table)
  ## calculate max difference and compare with tol
  diffs <- abs(prob_table$weights - weights)
  max_diff <- which(diffs == max(diffs))
  diff_value <- diffs[max_diff] / prob_table$weights[max_diff]
  if (diff_value > tol){ # if tolerance is not reached, continue
    message("Iteration #", iter, ". difference = ", diff_value)
    iter <- iter + 1
    prob_table$weights <- weights
  } else {
    message("Iteration #", iter, ". difference = ", diff_value, ". Stopping")
    prob_table$weights <- weights
    break
  }
}
```

```
## Iteration #1. difference = 1.20406022416932
## Iteration #2. difference = 2.33367775076681e-16. Stopping
```

```
## The marginal probabilities for d1 and d2 are now within the tolerance bounds:
prob_table
```

```
##    A  B    weights     d2_pA     d2_pB     d1_pA     d1_pB
## 1 a1 b1 0.07100592 0.3076923 0.2307692 0.3076923 0.2307692
## 2 a2 b1 0.11893491 0.5153846 0.2307692 0.5153846 0.2307692
## 3 a3 b1 0.04082840 0.1769231 0.2307692 0.1769231 0.2307692
## 4 a1 b2 0.08047337 0.3076923 0.2615385 0.3076923 0.2615385
## 5 a2 b2 0.13479290 0.5153846 0.2615385 0.5153846 0.2615385
## 6 a3 b2 0.04627219 0.1769231 0.2615385 0.1769231 0.2615385
## 7 a1 b3 0.15621302 0.3076923 0.5076923 0.3076923 0.5076923
## 8 a2 b3 0.26165680 0.5153846 0.5076923 0.5153846 0.5076923
## 9 a3 b3 0.08982249 0.1769231 0.5076923 0.1769231 0.5076923
```

```
## The weights can then be used to adjust the probabilities in a weighted bootstrapping
## algorithm, by assigning the weight matching the combination of A and B for each individual.
```

References

Fienberg, Stephen E. 1970. "An Iterative Procedure for Estimation in Contingency Tables." *The Annals of Mathematical Statistics* 41 (3). Institute of Mathematical Statistics: pp. 907–17. http://www.jstor.org/stable/2239244.

Norman, P. 1999. "Putting Iterative Proportional Fitting on the Researcher's Desk." http://eprints.whiterose.ac.uk/5029/.

Speed, Terry P. 2005. "Iterative Proportional Fitting." In *Encyclopedia of Biostatistics*. John Wiley & Sons, Ltd. doi:10.1002/0470011815.b2a10027. http://dx.doi.org/10.1002/0470011815.b2a10027.

STROBE 2007 (v4) checklist of items to be included in reports of observational studies in epidemiology*

Checklist for cohort, case-control, and cross-sectional studies (combined)

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	3-4
Objectives	3	State specific objectives, including any pre-specified hypotheses	3-4
Methods			
Study design	4	Present key elements of study design early in the paper	4-5
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	5
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	5 NA NA
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case	6 NA
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	5-6
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	5
Bias	9	Describe any efforts to address potential sources of bias	5-6
Study size	10	Explain how the study size was arrived at	5
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	5-6
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	5-6
		(*b*) Describe any methods used to examine subgroups and interactions	5
		(*c*) Explain how missing data were addressed	5
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed	5

		Cross-sectional study—If applicable, describe analytical methods taking account of sampling strategy	
		(*e*) Describe any sensitivity analyses	5-6
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	6
		(b) Give reasons for non-participation at each stage	NA
		(c) Consider use of a flow diagram	See cover letter
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	6 and table 1
		(b) Indicate number of participants with missing data for each variable of interest	Table 1
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	6
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	6, table 2 and figure 1
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	NA
		Cross-sectional study—Report numbers of outcome events or summary measures	NA
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	6-7, tables 3 and 4
		(*b*) Report category boundaries when continuous variables were categorized	5
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	NA
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	6-7
Discussion			
Key results	18	Summarise key results with reference to study objectives	7
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	8-9
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	9
Generalisability	21	Discuss the generalisability (external validity) of the study results	8-9
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	17

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.