

Use of primary care electronic medical record database in drug efficacy research on cardiovascular outcomes: comparison of database and randomised controlled trial findings

Richard L Tannen, professor of medicine, Mark G Weiner, associate professor of medicine, Dawei Xie, assistant professor of biostatistics and epidemiology

University of Pennsylvania School of Medicine, 295 John Morgan Building, 36th and Hamilton Walk, Philadelphia, PA 19104, USA

Correspondence to: R L Tannen tannen@mail.med.upenn.edu

Cite this as: *BMJ* 2009;338:b81
doi:10.1136/bmj.b81

ABSTRACT

Objectives To determine whether observational studies that use an electronic medical record database can provide valid results of therapeutic effectiveness and to develop new methods to enhance validity.

Design Data from the UK general practice research database (GPRD) were used to replicate previously performed randomised controlled trials, to the extent that was feasible aside from randomisation.

Studies Six published randomised controlled trials.

Main outcome measure Cardiovascular outcomes analysed by hazard ratios calculated with standard biostatistical methods and a new analytical technique, prior event rate ratio (PERR) adjustment.

Results In nine of 17 outcome comparisons, there were no significant differences between results of randomised controlled trials and database studies analysed using standard biostatistical methods or PERR analysis. In eight comparisons, Cox adjusted hazard ratios in the database differed significantly from the results of the randomised controlled trials, suggesting unmeasured confounding. In seven of these eight, PERR adjusted hazard ratios differed significantly from Cox adjusted hazard ratios, whereas in five they didn't differ significantly, and in three were more similar to the hazard ratio from the randomised controlled trial, yielding PERR results more similar to the randomised controlled trial than Cox ($P < 0.05$).

Conclusions Although observational studies using databases are subject to unmeasured confounding, our new analytical technique (PERR), applied here to cardiovascular outcomes, worked well to identify and reduce the effects of such confounding. These results suggest that electronic medical record databases can be useful to investigate therapeutic effectiveness.

INTRODUCTION

The future widespread implementation of electronic records in clinical practice will provide an enormous opportunity for research related to medical treatments, provided this information is compiled into robust, well designed databases and analysed with appropriate

methods. By contrast, incorrect analyses could have important negative effects on medical treatment and health policy. Therefore, before implementation of this approach for assessing effectiveness of treatment, we need to assess the validity of the results from studies using such databases and of the study design and analytical strategies that are most likely to yield valid results. The need for further investigation into these strategies is widely supported.¹⁻⁷

Two major potential problems could arise in the use of medical record databases to provide reliable information concerning treatment outcomes: the quality of the data contained within the database and the ability of analyses of observational—that is, non-experimental—data to provide valid results.

Considerable controversy exists over whether observational studies can provide reliable information on effectiveness of therapeutics.¹²⁸⁻¹⁵ Because of their ability to balance measured and unmeasured confounders, randomised controlled trials remain the highest level of evidence, whereas the quality of evidence from observational studies is lower because of confounding by indication and other biases related to the effects of unmeasured covariates. Several comparative analyses suggest that observational studies often yield results reasonably consistent with those of randomised controlled trials. Nevertheless, there are several well documented examples where the results from observational studies were misleading.¹⁻³⁷⁸¹⁶¹⁷ Hormone treatment to protect against coronary artery disease in postmenopausal women is one highly cited example.⁷⁻²⁰ Some authorities believe that the results of observational studies should be an important component of evidence based medicine; some suggest their reliability is limited to conditions where confounding by indication is unlikely, as, for example, in studies of unanticipated adverse effects of drugs,²¹ whereas others are sceptical of their value.¹²⁷⁻¹⁵

An important limitation applicable to previous comparative analyses is that most of the observational studies did not have rigorous inclusion and exclusion

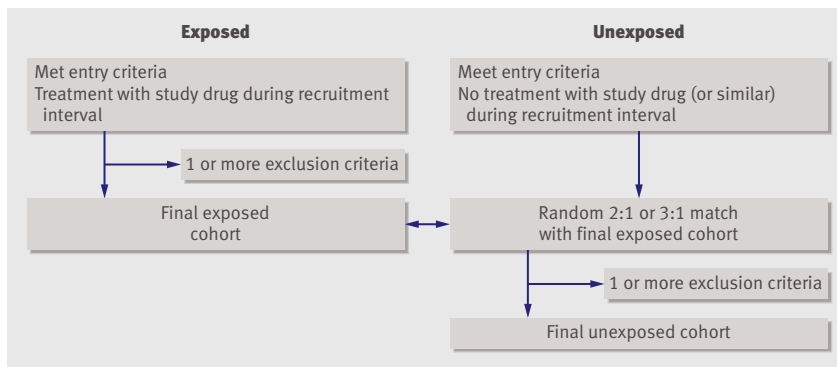


Fig 1 | Selection process for participants in database studies

criteria, exposure definitions, and outcomes identical to the randomised controlled trials so that lack of randomisation was not the only important difference.^{1 2 15 22 23}

To overcome these limitations in validating an observational study, we tested the value of a comprehensive longitudinal electronic clinical database, the UK General Practice Research Database (GPRD), using studies designed to replicate the design of previously performed randomised controlled trials to the extent that was feasible aside from randomisation.²⁴ Validity of the method was measured by comparing the outcomes of the replicated GPRD study with those of the randomised controlled trial.²⁵⁻²⁹ The GPRD study results depended on both the quality of the information in the database and whether observational data can reproduce results from a randomised controlled trial.

We examined both the potential research value of the electronic medical record database and the validity of observational studies. We also used a new analytical method, prior event rate ratio (PERR) adjustment, to enhance the validity of the results.

METHODS

GPRD database

The UK GPRD database contains information from the electronic medical records of primary care practices encompassing a representative sample of about 5.7% of the UK population during 1990-2000 and contains records of over eight million patients.^{24 25} It includes the complete primary care medical record, comprehensive information on essentially all medications prescribed, and information from outside consultants and admissions to hospital. The box details limitations and advantages of the database.

GPRD study protocol

Table 1 summarises database replications of six randomised controlled trials that have been performed and reported in detail elsewhere.^{20 30-34} As far as possible the database studies used the same inclusion and exclusion criteria, a similar study time frame, and a similar treatment regimen as the randomised trials.²⁵⁻²⁹ Thus the major primary difference was the lack of randomisation in the database studies, albeit that other

issues such as use of placebo, nature of healthcare delivery, and some characteristics of subjects entered into randomised trials compared with those existing in the general population can differ between a randomised trial and a database study.

Selection of the subjects for inclusion in the database studies followed the outline shown in figure 1. First the exposed cohort was selected from all database subjects who met the inclusion criteria and received treatment with the study treatment during a predefined recruitment interval. The exposed cohort was finalised after elimination of patients with exclusion criteria. Their start time was the day of the first prescription of the study drug. The unexposed cohort was selected from all patients who met the inclusion criteria but did not receive the study drug during the recruitment interval. They were then age and sex matched to the exposed patients with a computerised random selection program, and their start time was considered identical to that of the matched exposed patients. Then, those who had exclusion criteria were eliminated.

The selection process differed for the database matched to the Syst-Eur study because study entry and start time for both the exposed and unexposed cohorts was determined by measured blood pressure that indicated systolic hypertension.²⁵

All database studies ended on a predefined date or on outcome stop points defined in the randomised controlled trial. Patients were considered lost to follow-up if they left the practice or the practice was eliminated from the database before the end date. We analysed database studies using a simulated “intention to treat” paradigm where subsequent treatment of the exposed and unexposed patients did not modify study end time and also an “as treated” analysis in which the study ended for an exposed or unexposed patient who deviated from their treatment protocol.

Statistical analysis

We determined Cox unadjusted and adjusted hazard ratios for all outcomes. The adjusted hazard ratios used a predetermined set of potential confounders that included key demographics, medications at baseline, and identified medical conditions. We imputed missing values for systolic blood pressure, body mass index, and smoking³⁵ and created five separate datasets. The final estimates combined the results from the five datasets, as described previously.^{26 35-37}

We also analysed results with a propensity score approach, which used all demographics, drug use at baseline, and identified medical conditions as confounders.^{25 26 38-40} Propensity scores were estimated using logistic regression with the outcome being the indicator of treatment and the covariates being all confounders considered. For those with no missing data, all covariates were used; whereas for those with missing data (for body mass index, systolic blood pressure, or smoking), we used separate logistic regression models, which excluded the missing covariates, to estimate propensity scores. Analysis stratified by the propensity scores balances the treated and

untreated groups with respect to the observed covariates used in estimating the propensity scores. We determined outcome hazard ratios for each fifth of the propensity score and combined the five hazard ratios to determine an overall hazard ratio using a Cox model treating the fifths as strata with different baseline hazards. The propensity score thereby accounts for missing confounders in a different fashion from the multiple imputation method used with the Cox analysis. The matched database study for Syst-Eur, our first study, was analysed only with propensity score analysis.²⁵

We also used a prior event rate ratio (PERR) approach to adjust the Cox hazard ratio, as described recently.^{28,29} This analysis requires that neither the exposed nor unexposed patients are treated with the study drug before the start of the study. It assumes that the hazard ratio of the exposed to unexposed for a specific outcome before the start of the study reflects the combined effect of all confounders (both measured and unmeasured) independent of any influence of treatment.

To apply the PERR adjustment method, we divided the unadjusted hazard ratio of exposed versus unexposed groups during the study by the unadjusted hazard ratio of exposed versus unexposed “before” the study. Thus if p =prior events and s =study events, the calculation is: PERR adjusted HR=HRs/HRp. We obtained confidence intervals for the PERR adjusted hazard ratio using a bootstrap technique.²⁸ Hazard ratios are reported because of variable observation times for patients both before and during the study; though incidence rate ratios produced similar results.

In all studies we carried out the PERR analysis using a subset of patients who did not take the study drug at any time before the start of the study. In no instance did

Cox adjusted hazard ratios for this subset differ meaningfully from the results in the overall cohort.

The time interval used to assess previous events encompassed 1 January 1987 to the patient's start time. If a patient had no medical or treatment record before that date, their time interval began on the earliest subsequent date with a record. If they had no records before the study start time, they were not used in this analysis. The average time of the previous period for all the outcomes assessed averaged 3.52 years (range 2.8-3.9 years). Analysis of the impact of the duration of the previous time period using the empirical data in these studies suggested that encompassing events from 3-4.5 years before study start time did not meaningfully influence the results of the PERR analysis.

We compared differences between the hazard ratio from the randomised trial and the database using a standard normal z test, where the z score was obtained from the difference between the logarithm of the hazard ratio divided by the standard error of that difference.²⁵

RESULTS

We collated and analysed the collective results of six database studies reported previously.²⁵⁻²⁹

Comparability between replicated database study and randomised controlled trial

The size of the unexposed group in the database study was always larger than the placebo group of the randomised controlled trials (table 1). The exposed group in the database study, however, was smaller than the treated cohort in half the randomised controlled trials. Furthermore, the database was inadequate to replicate several randomised controlled trials because of an insufficient number of exposed patients.

Table 1 | Comparison of study characteristics in randomised controlled trials and general practice research database (GPRD)

Study	No of subjects				Treatment protocol	
	Trials		Database		Trials	Database
Syst-Eur*	2398	2297	2815	13 956	Nitrendipine, enalapril or hydrochlorothiazide. Target: 20 mm Hg decrease SBP	Dihydropyridine calcium channel blocker, ACE inhibitor, or thiazide diuretic. No target
WHI (Women's Health Initiative):						
Intact uterus†	8506	8902	13 658	37 730	Conjugated oestrogen 0.625 mg/day; medroxyprogesterone 5.0 mg/day	Conjugated oestrogen 0.625 mg/day; norgestrel 150 µg on days 17-28
Hysterectomy‡	5310	5429	6890	11 572	Conjugated oestrogen 0.625 mg/day	Conjugated oestrogen 0.625 mg/day
4S§	2221	2223	1280	2871	Simvastatin 20 mg/day. Target: total cholesterol 115-200 mg/dl	Any statin (80% received simvastatin). No target
HOPE¶	4645	4652	2812	26 286	Ramipril 10 mg/day	Any ACE I inhibitor with ramipril equivalent >4 mg/day (average dose 6.8 mg/day)
EUROPA**	6110	6108	2668	12 705	Perindopril 8 mg/day (ramipril equivalent 10 mg/day)	Any ACE inhibitor with ramipril equivalent >4 mg/day (average dose 6.5 mg/day)

ACE=angiotensin converting enzyme.

*Antihypertensive treatment of older patients with isolated systolic hypertension in Europe.^{25 30}

†Postmenopausal women treated with combined hormone replacement.^{20 26 41}

‡Postmenopausal women with previous hysterectomy treated solely with conjugated oestrogen.^{27 31 41}

§Scandinavian simvastatin survival study of hypercholesterolaemic subjects with coronary artery disease treated with simvastatin.^{28 32}

¶Heart outcomes prevention evaluation study of ramipril (angiotensin converting enzyme inhibitor), treatment of patients either with established or at high risk for coronary artery disease.^{29 33}

**European trial on reduction of cardiac events with perindopril (angiotensin converting enzyme inhibitor) in patients with stable coronary artery disease.^{29 34}

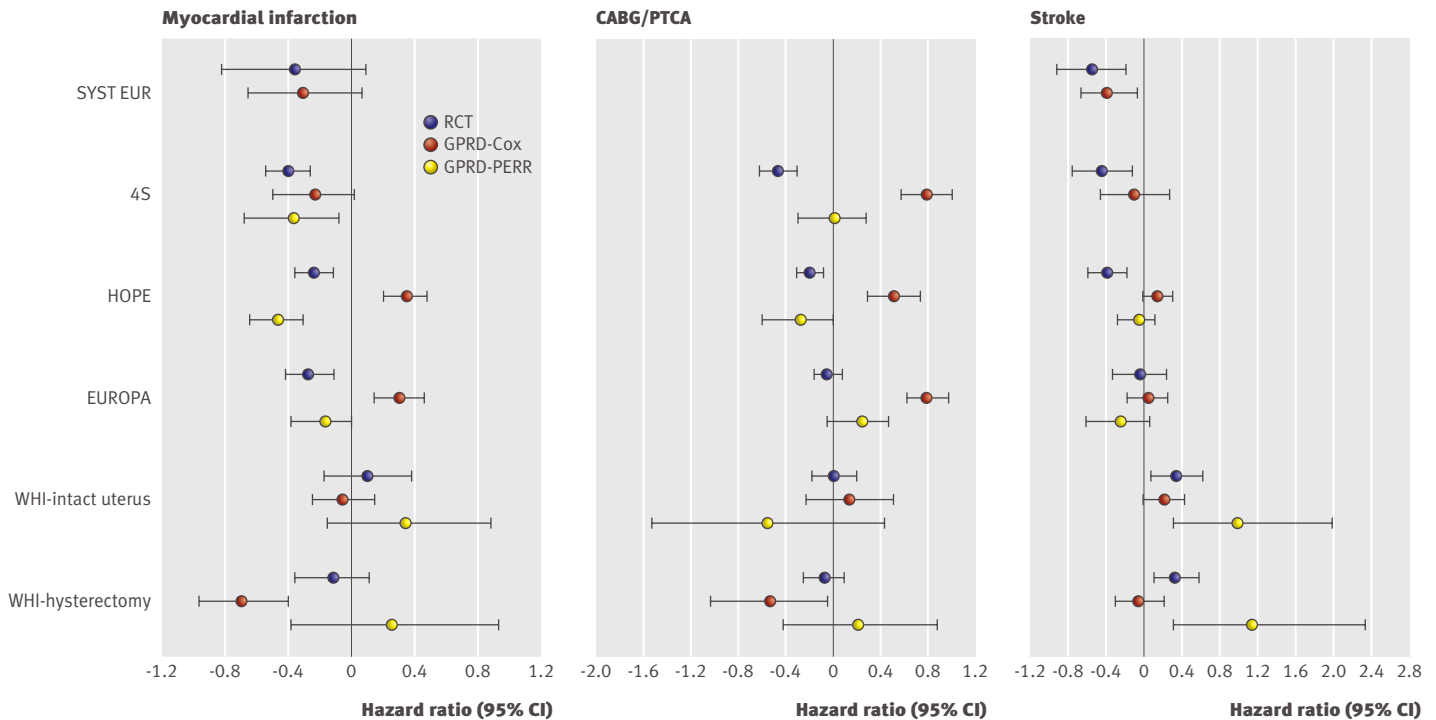


Fig 2 | Comparisons between hazard ratios from randomised controlled trials (RCT) and adjusted hazard ratios for respective database studies. Data plotted as natural logarithms, so 0 on x axis indicates no difference between exposed and unexposed cohort. Database adjusted hazard ratios shown with both Cox and prior event rate ratio (PERR) adjustment analysis. Results are shown for myocardial infarction, stroke, and coronary revascularisation (CABG/PTCA). GPRD=general practice research database

Although entry criteria were similar for the database studies and randomised trials, the database cohort typically differed from the respective trials in their baseline demographic characteristics, existing comorbidities, and use of cardiovascular drugs.²⁵⁻²⁹

The database treatment protocol precisely replicated the trial in only one study (WHI-hysterectomy (see table 1). The other database studies used the same class of drug, rather than specific drug used in the trial. Furthermore, identical dosing regimens could not be replicated. It is worth noting that the prescription database in GPRD can actually track data on medication prescribing better than many randomised controlled trials.

Finally in contrast with the randomised controlled trials, where randomisation resulted in similar baseline health profiles of the treated and placebo arms, all the database studies except Syst-Eur exhibited differences in the baseline characteristics of the exposed and unexposed groups.

Comparison of outcomes in the database studies and randomised controlled trials

We focused on randomised controlled trials with primary cardiovascular outcomes because they could be replicated reasonably without the need for laboratory data. We report on death, myocardial infarction, stroke, and coronary revascularisation (coronary artery bypass grafts or percutaneous transluminal coronary angioplasty (CABG/PTCA) or both). These cardiovascular outcomes should be the least

susceptible to misclassification errors. Other outcome results are provided in the primary publications, and the results for breast cancer, colon cancer, and hip fracture were similar in both the “intact uterus” and “hysterectomy” WHI randomised controlled trials and their respective database studies.²⁵⁻²⁹

Table 2 and figure 2 show cardiovascular outcomes and statistical comparisons for the six database studies and trials. We have shown simulated “intention to treat” results, but results of the “as treated” analyses did not differ meaningfully. Cox adjusted and prior event rate ratio (PERR) adjusted hazard ratios (performed in five studies) are also shown.

Propensity score analyses (table 3) did not differ meaningfully from the analysis with Cox adjusted hazard ratios. A minor exception was the death outcomes in the HOPE and EUROPA studies, where the propensity score adjusted hazard ratios were slightly lower than the Cox adjusted hazard ratios and slightly more similar to the hazard ratios from the randomised controlled trial.

Results from the WHI randomised controlled trial for the entire cohort and also subdivided by age were reported.^{20,31-41} We compared the database studies to the overall WHI randomised controlled trial and also to the results restricted to women aged <70, an age profile more comparable with the study cohorts in the database.

Cardiovascular outcomes

In nine of 17 comparisons of cardiovascular outcomes there was no significant difference between the Cox

adjusted hazard ratios from the database and the hazard ratios from the randomised controlled trials (see table 2, which compares the trial hazard ratio, the database Cox adjusted hazard ratio, and the database-PERR adjusted hazard ratios). In none of these nine comparisons did the PERR analysis differ significantly from either the trial hazard ratios or the Cox adjusted hazard ratios.

In eight of the 17 comparisons, however, the Cox adjusted hazard ratios differed significantly from the trial hazard ratios, suggesting the presence of unmeasured confounding. In seven of these eight instances the PERR adjusted hazard ratios differed significantly from the Cox adjusted hazard ratios, and either did not differ significantly (five outcomes) or were more similar (two outcomes) to the trial hazard ratio. In the other outcome the PERR hazard ratio was more similar to the trial but did not differ significantly from the Cox adjusted hazard ratio. A Wilcoxon signed rank test showed that when the Cox adjusted hazard ratio differed significantly from the trial hazard ratio ($n=8$), the PERR adjusted hazard ratio was significantly ($P<0.05$) more similar to the trial hazard ratio than the Cox adjusted hazard ratio.

As the 17 outcomes analysed came from six studies, it is reasonable to question the analysis of each outcome as an independent data point. As shown in the 4S study, however, the two outcomes (myocardial infarction and coronary revascularisation) clearly behaved independently of one another. Unmeasured confounding affected revascularisation but had no discernible effect on myocardial infarction. As the PERR analysis is outcome specific and derived entirely in that fashion, it seems reasonable to analyse the data assuming that each individual outcome is independent.

In the aggregate, when the outcome results from the database studies analysed by conventional statistical methods are confirmed or corrected by the PERR method, they are largely comparable with the results from the respective randomised controlled trials.

The large confidence intervals in the PERR analysis of all the WHI outcomes, which limits the interpretation of this data, were due to the small number of previous events. Surprisingly, despite this limitation,

the PERR adjusted hazard ratio was significantly higher than the Cox adjusted hazard ratio and not different from the randomised controlled trial hazard ratio for both the myocardial infarction and stroke outcomes in the WHI-hysterectomy study, suggesting the presence of unmeasured confounding. The likelihood that unmeasured confounding influenced these two outcomes is consistent with the significant difference between the Cox adjusted hazard ratios and the randomised controlled trial hazard ratios.

Death

We have shown only Cox adjusted hazard ratios for death because PERR adjustment cannot be done. The Cox and the propensity score adjusted hazard ratios for death resembled the randomised controlled trial results in three studies; however, they were higher than the trial in the Syst-Eur study and lower in both the WHI studies. The WHI results on death should be interpreted cautiously because in both studies a subset of the overall cohort that was not missing any data on baseline body mass index, systolic blood pressure, or smoking did not show a significant decrease in death, despite results comparable with the overall cohort for all other outcomes.

DISCUSSION

Despite its shortcomings, this careful, albeit not exhaustive, comparison between randomised controlled trials and observational studies using data from an electronic primary care medical record database reveals several important insights. From an overall perspective, our results suggest that observational studies using databases might produce valid results concerning the efficacy of cardiovascular drug treatments.

Rigour of database studies

Our studies comparing performance of the database and randomised controlled trials were performed in as rigorous fashion as possible.

In addition to using similar inclusion and exclusion criteria and relatively similar time frames, we analysed studies with both a simulated “intention to treat” and an “as treated” design. We analysed data with multiple imputation plus Cox adjusted hazard ratios, and also propensity score plus stratified Cox unadjusted hazard ratios. The propensity score is useful to identify heterogeneity and also incorporates missing data into the analysis in a fashion different from the multiple imputations used with the primary Cox method. We used a subset of the overall cohort without “missing data” on the key confounders (systolic blood pressure, body mass index, and smoking) as a secondary verification analysis to ensure that missing data did not influence the results in the overall cohort. We assessed use of non-study drugs to confirm that cointervention during the study did not account for the results. Computerised random matching and thereby start time delineation for the unexposed

The GPRD database

Advantages

- Comprehensive national healthcare system
- Representative sample of entire population
- All care centralised in general practice record
- All medications prescribed by general practitioner, generated by computer
- Contains around eight million patients

Limitations

- Lacks direct link to laboratory data (laboratory data inadequate)
- Missing data on smoking, systolic blood pressure, body mass index (about 30%)
- Limited data on onset of menopause
- Limited data on admission to hospital
- Lacks direct link to death certificates (cause of death not reliable)

group obviated the potential for unanticipated bias related to start time in the unexposed group.

Overall study results

We analysed results of the outcomes for myocardial infarction, stroke, coronary revascularisation, and death for six comparative studies (table 2 and fig 2). We examined the aggregate database study results with conventional biostatistical analyses (Cox adjusted hazard ratios or propensity score analyses, or both) and our newly described prior event rate ratio (PERR) adjustment technique.^{28 29}

When analysed with conventional biostatistical analyses, the database outcome results (independent of death) did not differ significantly from those in the randomised controlled trial in nine of the 17 comparisons. In no instance did the PERR analysis differ significantly from the randomised controlled trial, when there was no difference between the conventional analyses and the trial.

As shown in table 2 and figure 2, when the database outcomes analysed with conventional biostatistical techniques differed significantly from the trial, the PERR analysis results were either not significantly different from or much more similar to the trial results.

The instances where the database results analysed by conventional biostatistical methods differed importantly from the results in the trial presumably reflect unmeasured confounding by indication in the database studies. Thus our findings support concerns that the validity of observational studies must always be viewed with circumspection. The studies reported herein, however, suggest that the PERR technique can identify (by differing from the results with standard statistical methods) and largely correct for the effects of unmeasured confounding, when it exists. The availability in the database of previous event rates, rather than only prevalence data, permitted performance of this analysis.

PERR analytical technique

The underlying hypothesis of the PERR analytical technique is that a comparison between the event rate for a specific outcome in a cohort's exposed and unexposed patients before entry into the study should reflect the effect of all confounders on that specific outcome independent of the effect of treatment. This assumption holds only when neither the exposed nor unexposed patients have been treated with the study drug before the start of the study. If so, the ratio between the previous events in the exposed and

Table 2 | Comparison of outcome hazard ratios in randomised controlled trials and general practice research database (GPRD)

	Death	Myocardial infarction	Stroke	CABG/PTCA
Syst-Eur				
Trial	0.86 (0.67 to 1.09)	0.70 (0.44 to 1.09)	0.58 (0.40 to 0.83)	—
GPRD-Cox	1.23 (1.00 to 1.50)*	0.74 (0.52 to 1.07)	0.68 (0.51 to 0.94)	—
WHI-intact uterus				
Trial	0.98 (0.82 to 1.15)	1.11 (0.84 to 1.47)†	1.41 (1.07 to 1.85)	1.01 (0.83 to 1.22)
GPRD-Cox	0.75 (0.65 to 0.86)*	0.95 (0.78 to 1.16)	1.23 (0.99 to 1.52)	1.15 (0.79 to 1.67)
GPRD-PERR	—	1.40 (0.87 to 2.44)	2.63 (1.38 to 7.43)	0.57 (0.22 to 1.56)
GPRD-no missing‡	0.91 (0.79 to 1.05)	—	—	—
WHI-hysterectomy				
Trial	1.01 (0.88 to 1.22)	0.89 (0.70 to 1.12)	1.39 (1.10 to 1.77)	0.93 (0.78 to 1.10)
GPRD-Cox	0.68 (0.57 to 0.81)*	0.50 (0.38 to 0.67)*	0.95 (0.74 to 1.23)*	0.59 (0.36 to 0.95)
GPRD-PERR	—	1.28 (0.69 to 2.56)§	3.06 (1.39 to 10.31)§	1.22 (0.67 to 2.42)
GPRD-no missing‡	0.82 (0.66 to 1.02)	—	—	—
4S				
Trial	0.70 (0.58 to 0.85)	0.67 (0.58 to 0.77)	0.64 (0.47 to 0.88)	0.63 (0.54 to 0.74)
GPRD-Cox	0.71 (0.53 to 0.96)	0.79 (0.61 to 1.02)	0.90 (0.63 to 1.30)	2.22 (1.80 to 2.75)*
GPRD-PERR	—	0.69 (0.51 to 0.93)	NA¶	1.00 (0.75 to 1.33)*§
HOPE				
Trial	0.84 (0.75 to 0.95)	0.79 (0.70 to 0.89)	0.68 (0.56 to .84)	0.82 (0.74 to 0.92)
GPRD-Cox	0.94 (0.85 to 1.03)	1.42 (1.23 to 1.61)*	1.16 (0.99 to 1.35)*	1.67 (1.34 to 2.07)*
GPRD-PERR	—	0.62 (0.53 to 0.74)*§	0.94 (0.77 to 1.14)*	0.75 (0.56 to 1.01)§
EUROPA				
Trial	0.89 (0.77 to 1.02)	0.76 (0.66 to 0.89)	0.96 (0.72 to 1.28)	0.96 (0.85 to 1.08)
GPRD-Cox	1.06 (0.95 to 1.19)	1.36 (1.16 to 1.58)*	1.04 (0.84 to 1.29)	2.20 (1.85 to 2.62)*
GPRD-PERR	—	0.84 (0.69 to 1.01)§	0.77 (0.55 to 1.07)	1.26 (.97 to 1.62)§

CABG/PTCA=coronary artery bypass grafts or percutaneous transluminal coronary angioplasty

*Significant difference (P<0.05) compared with trial.

†Trial values for myocardial infarction reflect WHI re-analysis by age, encompassing 50-70 years.

‡Subset not missing any data for BMI, systolic blood pressure, or smoking.

§Significant difference (P<0.05) compared with GPRD Cox adjusted hazard ratio.

¶PERR could not be done because stroke was study exclusion criteria.

unexposed patients should reflect the aggregate effect of all identified and unidentified confounders.

Therefore, when the unadjusted incidence rate ratio or hazard ratio of that outcome during the study is divided by the ratio for that outcome before the study, this adjustment should correct for the aggregate effects of all identified and unmeasured confounders.

When there are no unmeasured confounders, reflected by similar results of the database Cox adjusted hazard ratio and the randomised controlled trial hazard ratio, the PERR adjusted results should be similar to the Cox adjusted hazard ratio. Based on the empirical findings in these studies, the PERR adjustment seemed to function in this fashion.

When there are unmeasured confounders, presumably resulting from confounding by indication, the results of the PERR adjusted hazard ratio and the Cox adjusted hazard ratio should differ. Our empirical results show that in every instance where the comparison of the Cox adjusted hazard ratio in the database study differed from the results of the trial, suggesting the presence of “unidentified confounding,” the PERR adjustment yielded a result much more consistent with the findings in the trial. Of most importance in all but one instance where unmeasured confounding seemed to be present, the PERR adjusted value identified the presence of unmeasured confounding by differing significantly from the Cox adjusted hazard ratio.

Identification of the PERR method emerged from these studies because the direct comparison of the database observational study and the randomised controlled trial provided a presumed correct answer against which to validate the database results. Further investigation is necessary to fully validate the PERR technique. More extensive statistical simulation studies would determine its limitations and applications and the applicability of the method to additional outcomes. It is also important to appreciate that this technique is outcome specific; it cannot be extrapolated from one

outcome to another. Finally, it is restricted to outcomes for which previous events can be ascertained. If an outcome was a study exclusion criterion, it cannot be analysed with this approach, nor can it be applied to death.

The PERR method differs and seems to be more widely applicable than other methods that have been developed in an attempt to address hidden bias.⁴² As confirmed in our studies, propensity score analysis does not overcome unmeasured confounding. When combined with sensitivity analyses, however, it might provide results that can be interpreted as unlikely to have been influenced by unmeasured covariates.⁴³⁻⁴⁵ Recently, propensity scores combined with regression calibration were used to address unobserved variables under certain conditions.⁴⁶⁻⁴⁷

Instrumental variable analysis, used commonly in economics, has also been used to address unmeasured confounding. An instrumental variable analysis requires identification of a factor that affects the assignment to treatment but has no direct effect on the outcome.⁴⁸⁻⁵⁰ Its applicability and validity for studies of therapeutic efficacy have not been widely examined.⁴²⁻⁵¹⁻⁵² Some have suggested that this technique is most suited to address health policy issues rather than specific clinical issues of treatment effectiveness.⁴⁸

Both the propensity score calibration and the instrumental variable analysis methods have important constraints. The propensity score calibration technique requires the presence of a validation study, whereas the instrumental variable analysis requires identification of an appropriate instrument. These requirements limit their applicability to a wide variety of studies.

Of interest, the DID (difference-in-differences) method used in economic studies, has some similarities to the PERR method in that it compares the differences between the difference in before and after behaviour in two groups.⁵³⁻⁵⁵ The key assumption behind the DID method, similar to PERR, is that the distribution of the

Table 3 | Hazard ratios adjusted for Cox analysis and propensity scores in general practice research database (GPRD)

	Death	Myocardial infarction	Stroke	CABG/PTCA
WHI-intact uterus				
Cox	0.75 (0.65 to 0.86)	0.95 (0.78 to 1.16)	1.23 (0.99 to 1.52)	1.15 (0.79 to 1.67)
Propensity score	0.78 (0.69 to 0.89)	0.97 (0.80 to 1.17)	1.16 (0.95 to 1.42)	1.17 (0.82 to 1.70)
WHI-hysterectomy				
Cox	0.68 (0.57 to 0.81)	0.50 (0.38 to 0.67)	0.95 (0.74 to 1.23)	0.59 (0.36 to 0.95)
Propensity score	0.67 (0.57 to 0.80)	0.48 (0.36 to 0.64)	0.93 (0.72 to 1.19)	0.59 (0.37 to 0.96)
4S				
Cox	0.71 (0.53 to 0.96)	0.79 (0.61 to 1.02)	0.90 (0.63 to 1.30)	2.22 (1.80 to 2.75)
Propensity score	0.65 (0.48 to 0.89)	0.76 (0.59 to 1.00)	0.87 (0.60 to 1.27)	1.93 (1.55 to 2.40)
HOPE				
Cox	0.94 (0.85 to 1.03)	1.42 (1.23 to 1.61)	1.16 (0.99 to 1.35)	1.67 (1.34 to 2.07)
Propensity score	0.79 (0.72 to 0.88)*	1.36 (1.17 to 1.58)	1.04 (0.89 to 1.23)	1.55 (1.24 to 1.94)
EUROPA				
Cox	1.06 (0.95 to 1.19)	1.36 (1.16 to 1.58)	1.04 (0.84 to 1.29)	2.20 (1.85 to 2.62)
Propensity score	0.84 (0.75 to 0.95)*	1.31 (1.12 to 1.53)	0.86 (0.69 to 1.08)	2.37 (1.98 to 2.85)

CABG/PTCA=coronary artery bypass grafts or percutaneous transluminal coronary angioplasty

*Significant difference (P<0.05) compared with Cox adjusted hazard ratio.

WHAT IS ALREADY KNOWN ON THIS TOPIC

Two major potential problems could impede the capability of an electronic medical record database to provide reliable information concerning drug efficacy: the quality of the data contained within the database and the ability of analyses of observational—that is, non-experimental—data to provide valid results

The quality of evidence from observational studies is less than from randomised controlled trials because of confounding by indication and other biases related to the effects of unmeasured covariates

WHAT THIS STUDY ADDS

Although observational studies are subject to unmeasured confounding, a new analytical technique, prior event rate ratio (PERR) adjustment, can identify and reduce unmeasured confounding

Data from properly constructed electronic medical record databases, when analysed with standard statistical methods along with the PERR method, can reveal important insights into the efficacy of medical treatment

unobserved confounding variables in the treated group and the comparison group and the effect of these unobserved confounding variables on the outcome remains the same before and during the study period. The DID method is also used commonly in psychology, where it is called the before and after design with an untreated comparison group.^{56,57}

Death was significantly higher in one of our database studies (Syst-Eur) and it seemed to be significantly lower in both of the database comparisons with the WHI randomised controlled trial; however, for the reasons enumerated these latter results should be interpreted cautiously.

Future perspective and study limitations

Thus it seems from our studies that an electronic medical record database can be an important tool for ascertaining evidence based decisions with regard to treatment. To maximise the value of future databases they should be designed with all the advantages enumerated for GPRD and also should overcome its limitations (see box). Ideally future databases should be much larger than GPRD, which includes about eight million patients. On the basis of our work to date, we estimate that 40-50 million patients are needed for the breadth of future studies we can envisage.

Studies using such databases would not replace the need to do randomised controlled trials but could serve as an important tool to supplement the contributions of trials to evidence based medicine. One example among many is to generalise the results of randomised controlled trials. Although we have not comprehensively examined this issue, our studies have shown the feasibility of further generalising the results of the Syst-Eur and WHI randomised controlled trials.^{25,58,59}

As well as the need for further validation of the PERR technique, several other limitations apply to this investigative effort. The PERR technique should be viewed currently as applicable only to analysis of a study using a design similar to ours, which includes similar inclusion and exclusion criteria for the exposed and unexposed and a defined study start, recruitment

interval, and end time. Furthermore, the random matching technique might be critical to assure that bias does not exist in the start time for unexposed patients. Application of the PERR technique to other study designs will require its validation under those conditions.

Another potential shortcoming of our studies is the inability to exactly replicate all aspects of the randomised controlled trial independent of randomisation, such as exact dose of study drug, the role of placebos, the possibilities of differences in health care, and other differences between participants entered in randomised controlled trials and those in the general population. In addition, there is also the possibility of inaccuracy of information in the database (for instance, misclassification of outcome, ascertainment bias, etc). The reasonably similar results of the database studies and comparative randomised controlled trials, however, suggest these were not major problems.

Our current view is that the PERR analysis should not be performed in isolation. We would recommend its use along with conventional biostatistical analyses. When the conventional and PERR analyses are similar, “unmeasured confounding” would seem unlikely; whereas when they differ “unmeasured confounding” would seem likely. When unmeasured confounding seems to be present, the PERR analysis seems to yield a more valid result, but additional evaluation is required to ascertain the veracity of this suggestion.

Contributors: RLT and MHW contributed to conception and design; analysis and interpretation of data; drafting and revision of article; and final approval of published version. CX contributed to design, analysis and interpretation of data, drafting and revision of article, and final approval of published version.

We gratefully acknowledge the assistance of Xingmei Wang, who assisted with the biostatistical analyses, and James Lewis and Stephen Kimmel for their insightful review of this manuscript.

Contributors: RLT and MHW contributed to conception and design; analysis and interpretation of data; drafting and revision of article; and final approval of published version. CX contributed to design, analysis and interpretation of data, drafting and revision of article, and final approval of published version.

Funding: This work was supported by the National Institutes of Health research grant R01-HL 073911.

Competing interests: None declared.

Ethical approval: Not required.

Provenance and peer review: Not commissioned; externally peer reviewed.

- 1 Britton A, McKee M, Black N, McPherson K, Sanderson C, Bain C. Choosing between randomised and non-randomised studies: a systematic review. *Health Technol Assess* 1998;2:1-124.
- 2 MacLehose RR, Reeves BC, Harvey IM, Sheldon TA, Russell IT, Black AMS. A systematic review of comparisons of effect-sizes derived from randomized and non-randomised studies. *Health Technol Assess* 2000;4:1-154.
- 3 Benson K, Hartz AJ. A comparison of observational studies and randomized controlled trials. *N Engl J Med* 2000;342:1878-86.
- 4 Concato J, Horwitz RJ. Beyond randomized versus observational studies. *Lancet* 2004;363:1660-1.
- 5 Sacks HS. Observational studies and randomized trials. *N Engl J Med* 2000;343:1195.
- 6 Ioannidis JPA, Haidich AB, Lau J. Any casualties in the clash of randomised and observational evidence? *BMJ* 2001;322:879-80.
- 7 Avorn J. In defense of pharmacoepidemiology—embracing the yin and yang of drug research. *N Engl J Med* 2007;357:2219-21.

- 8 Ioannidis JPA, Haidich AB, Pappa M, Pantazis N, Kokori SI, Tektonidou MG, et al. Comparison of evidence of treatment effects in randomized and nonrandomized studies. *JAMA* 2001;286:821-30.
- 9 Black N. Why we need observational studies to evaluate the effectiveness of health care. *BMJ* 1996;312:1215-8.
- 10 Kunz R, Oxman AD. The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. *BMJ* 1998;317:1185-90.
- 11 Pocock SJ, Elbourne DR. Randomized trials or observational tribulations? *N Engl J Med* 2000;342:1907-9.
- 12 McMahon AD. Observation and experiment with the efficacy of drugs: a warning example from a cohort of nonsteroidal anti-inflammatory and ulcer-healing drug users. *Am J Epidemiol* 2001;154:557-62.
- 13 MacMahon S, Collins R. Reliable assessment of the effects of treatment on mortality and major morbidity, II: observational studies. *Lancet* 2001;357:455-6.
- 14 Califf RM, Pryor DB, Greenfield J Jr. Beyond randomized clinical trials: applying clinical experience in the treatment of patients with coronary artery disease. *Circulation* 1986;74:1191-4.
- 15 Deeks JJ, Dinnes J, D'Amico RA, Sowden AJ, Sakaravitch C, Song F, Petticrew M, Altman DG. Evaluating non-randomised intervention studies. *Health Technol Assess* 2003;7:1-173.
- 16 McKee M, Britton A, Black N, McPherson K, Sanderson C, Bain C. Interpreting the evidence choosing between randomised and non-randomised studies. *BMJ* 1999;319:312-5.
- 17 Concato J, Shah N, Horwitz RL. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med* 2000;342:1887-92.
- 18 Grodstein F, Stampfer MJ, Manson JE, Colditz GA, Willett WC, Rosner B, et al. Postmenopausal estrogen and progestin use and the risk of cardiovascular disease. *J Engl J Med* 1996;335:453-61.
- 19 Grodstein F, Manson JE, Colditz GA, Willett WC, Speizer FE, Stampfer MJ. A prospective, observational study of postmenopausal hormone therapy and primary prevention of cardiovascular disease. *Ann Intern Med* 2000;133:933-41.
- 20 Writing Group for the Women's Health Initiative Investigators. Risks and benefits of estrogen plus progestin in health postmenopausal women. Principal results from the Women's Health Initiative randomized controlled trial. *JAMA* 2002;288:321-33.
- 21 Vandenberg JP. When are observational studies as credible as randomized trials? *Lancet* 2004;363:1728-31.
- 22 Coronary artery surgery study (CASS): a randomized trial of coronary bypass surgery. Comparability of entry characteristics and survival in randomized patients and nonrandomized patients meeting randomization criteria. *J Am Coll Cardiol* 1984;3:114-28.
- 23 Horwitz RL, Viscoli CM, Clemens JD, Sadock RT. Developing improved observational methods for evaluating therapeutic effectiveness. *Am J Med* 1990;89:630-8.
- 24 Gelfand JM, Margolis DJ, Dattani H. The UK general practice research database. In: Strom BL, ed. *Pharmacoepidemiology*. 4th ed. New York: John Wiley, 2005:337-46.
- 25 Tannen RL, Weiner MG, Marcus SM. Simulation of the Syst-Eur randomized control trial using a primary care electronic medical record was feasible. *J Clin Epidemiol* 2006;59:254-64.
- 26 Tannen RL, Weiner MG, Xie D, Barnhart K. A simulation using data from a primary care practice database closely replicated the women's health initiative trial. *J Clin Epidemiol* 2007;60:686-95.
- 27 Tannen RL, Weiner MG, Xie D, Barnhart K. Estrogen affects postmenopausal women differently than estrogen plus progestin therapy. *Hum Reprod* 2007;22:1769-77.
- 28 Weiner MG, Xie D, Tannen RL. Replication of the Scandinavian Simvastatin Survival Study using a primary care medical record database prompted exploration of a new method to address unmeasured confounding. *Pharmacoepidemiol Drug Saf* 2008;17:661-70.
- 29 Tannen RL, Weiner MG, Xie D. Replicated studies of two randomized trials of angiotensin converting enzyme inhibitors: further empirical validation of the "prior event rate ratio" to adjust for unmeasured confounding by indication. *Pharmacoepidemiol Drug Saf* 2008;17:671-85.
- 30 Staessen JA, Fagard R, Thijs L, Celis H, Arabidze GG, Birkenhager WH, et al. Randomised double-blind comparison of placebo and active treatment for older patients with isolated systolic hypertension. *Lancet* 1997;350:757-64.
- 31 Women's Health Initiative Steering Committee. Effects of conjugated equine estrogen in postmenopausal women with hysterectomy. The Women's Health Initiative randomized controlled trial. *JAMA* 2004;291:1701-12.
- 32 Scandinavian Simvastatin Survival Study Group. Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease: the Scandinavian Simvastatin survival study (4S). *Lancet* 1994;344:1383-9.
- 33 Heart Outcomes Prevention Evaluation Study Investigators. Effects of an angiotensin-converting-enzyme inhibitor, ramipril, on cardiovascular events in high risk patients. *N Engl J Med* 2000;342:145-53.
- 34 European trial on reduction of cardiac events with Perindopril in stable coronary artery disease investigators. Efficacy of perindopril in reduction of cardiovascular events among patients with stable coronary artery disease: randomized double-blind, placebo-controlled, multicentre trial (the EUROPA study). *Lancet* 2003;362:782-8.
- 35 Raghunathan TE, Solenberger PW, Van Hoewyk J. *IVAware: imputation and variance estimation software*. Ann Arbor, MI: Institute for Social Research, University of Michigan, 1998.
- 36 Rubin D. *Multiple imputation for non-response in surveys*. New York: John Wiley, 1987.
- 37 Raghunathan TE, Lepkowski JM, Van Hoewyk J, Solenberger PW. A multivariate technique for multiply imputing missing values using a sequence of regression models. *Surv Methodol* 2001;27:85-95.
- 38 Rosenbaum P, Rubin D. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983;70:41-55.
- 39 Rosenbaum P, Rubin D. Reducing bias in observational studies using subclassification on the propensity score. *J Am Stat Assoc* 1984;79:516-24.
- 40 D'Agostino RB, Rubin DB. Estimating and using propensity scores with partially missing data. *J Am Stat Assoc* 2000;95:749-59.
- 41 Rossouw JE, Prentice RL, Manson JE, Wu L, Barad D, Barnabei VM, et al. Postmenopausal hormone therapy and risk of cardiovascular disease by age and years since menopause. *JAMA* 2007;297:1465-77.
- 42 McMahon AD. Approaches to combat confounding by indication in observational studies of intended drug effects. *Pharmacoepidemiol Drug Saf* 2003;12:551-8.
- 43 Rosenbaum PR. Discussing hidden bias in observational studies. *Ann Intern Med* 1991;115:901-5.
- 44 Rubin DB. Estimating causal effects from large data sets using propensity scores. *Ann Intern Med* 1997;127:757-63.
- 45 Braitman LE, Rosenbaum PR. Rare outcomes, common treatments: analytic strategies using propensity scores. *Ann Intern Med* 2002;137:693-5.
- 46 Sturmer T, Schneeweiss, Rothman KJ, Avorn J, Glynn RJ. Performance of propensity score calibration. A simulation study. *Am J Epidemiol* 2007;165:1110-8.
- 47 Oakes JM, Church TR. Invited commentary: advancing propensity study methods in epidemiology. *Am J Epidemiol* 2007;165:1119-21.
- 48 McClellan M, McNeil BJ, Newhouse JP. Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality? Analysis using instrumental variables. *JAMA* 1994;272:859-66.
- 49 Newhouse JP, McClellan M. Econometrics in outcomes research: the use of instrumental variables. *Annu Rev Public Health* 1998;19:17-34.
- 50 Sukel TA, Fisher ES, Wennberg DE, Alter DA, Gottlieb DJ, Vermeulen MJ. Analysis of observational studies in the presence of treatment selection bias. Effect of invasive cardiac management on AMI survival using propensity score and instrumental variable methods. *JAMA* 2007;297:278-85.
- 51 Martens EP, Pestman WR, de Boer A, Belitser SV, Klungel OH. Instrumental variables. Applications and limitations. *Epidemiology* 2006;17:260-7.
- 52 Brookhart MA, Wang PS, Solomon DH, Schneeweiss S. Evaluating short-term drug effects using a physician-specific prescribing preference as an instrumental variable. *Epidemiology* 2006;17:268-75.
- 53 Gruber J. The incidence of mandated maternity benefits. *American Economic Review* 1994;84:622-41.
- 54 Meyer BD. Natural and quasi-experiments in economics. *J Business Econom Stat* 13:151-61.
- 55 Angrist JD, Krueger AB. Empirical strategies in labor economics. In: Ashenfelter O, Card D, eds. *Handbook of labor economics*. Amsterdam: Elsevier Science, 1999:1277-366.
- 56 Campbell D, Stanley J. *Experimental and quasi-experimental designs for research*. Boston, MA: Houghton-Mifflin, 1963.
- 57 Shadish WR, Cook TD, Campbell DT. *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin, 2002.
- 58 Weiner MG, Barnhart K, Xie D, Tannen RL. Hormone replacement therapy and coronary heart disease in young women. *Menopause* 2008;15:86-93.
- 59 Tannen RL, Weiner MG, Xie D, Barnhart K. Perspectives on hormone replacement therapy: the Women's Health Initiative and new observational studies sampling the overall population. *Fertil Steril* 2008;90:258-64.

Accepted: 19 October 2008