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

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.1-3 11-13 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.1-3 7-16 17 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.1-7 15

An important limitation applicable to previous comparative analyses is that most of the observational studies did not have rigorous inclusion and exclusion...
criteria, exposure definitions, and outcomes identical to the randomised controlled trials so that lack of randomisation was not the only important difference.² ¹² 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 Validation 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 ²⁹-³⁴ 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.  

We also used a prior event rate ratio (PERR) approach to adjust the Cox hazard ratio, as described recently. 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: \( \text{PERR adjusted HR} = \frac{\text{HR}_{\text{ex}}}{\text{HR}_{\text{un}}} \times \frac{\text{HR}_{\text{pre}}}{\text{HR}_{\text{pre}}^{\text{st}}} \). We obtained confidence intervals for the PERR adjusted hazard ratio using a bootstrap technique. 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 start study 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.

**RESULTS**

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

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)

<table>
<thead>
<tr>
<th>Study</th>
<th>Trials</th>
<th>Database</th>
<th>Treatment protocol</th>
<th>Database</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No of subjects</td>
<td></td>
<td>Nitrendipine, enalapril or hydrochlorothiazide.</td>
<td>Dihydropyridine calcium channel blocker, ACE inhibitor, or thiazide diuretic. No target</td>
</tr>
<tr>
<td></td>
<td>Treated</td>
<td>Placebo</td>
<td>Exposed</td>
<td>Unexposed</td>
</tr>
<tr>
<td>Syst-Eur*</td>
<td>2398</td>
<td>2297</td>
<td>2815</td>
<td>13 956</td>
</tr>
<tr>
<td>WHI (Women’s Health Initiative):</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intact uterus†</td>
<td>8506</td>
<td>8902</td>
<td>13 658</td>
<td>37 730</td>
</tr>
<tr>
<td>Hysterectomy‡</td>
<td>5310</td>
<td>5429</td>
<td>6890</td>
<td>11 572</td>
</tr>
<tr>
<td>4S§</td>
<td>2221</td>
<td>2223</td>
<td>1280</td>
<td>2871</td>
</tr>
<tr>
<td>HOPE¶</td>
<td>4645</td>
<td>4652</td>
<td>2812</td>
<td>26 286</td>
</tr>
<tr>
<td>EUROPA**</td>
<td>6110</td>
<td>6108</td>
<td>2668</td>
<td>12 705</td>
</tr>
</tbody>
</table>

ACE=angiotensin converting enzyme.

*Antihypertensive treatment of older patients with isolated systolic hypertension in Europe.**

†Postmenopausal women treated with combined hormone replacement.

‡Postmenopausal women with previous hysterectomy treated solely with conjugated oestrogen.

§Scandinavian simvastatin survival study of hypercholesterolaemic subjects with coronary artery disease treated with simvastatin.

¶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.

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

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. 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
Limitations

Advantages

The GPRD database

**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.

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Comparison of outcome hazard ratios in randomised controlled trials and general practice research database (GPRD)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Death</td>
</tr>
<tr>
<td>Syst-Eur</td>
<td></td>
</tr>
<tr>
<td>Trial</td>
<td>0.86 (0.67 to 1.09)</td>
</tr>
<tr>
<td>GPRD-Cox</td>
<td>1.23 (1.00 to 1.50)*</td>
</tr>
<tr>
<td>WHI-intact uterus</td>
<td>0.98 (0.82 to 1.15)</td>
</tr>
<tr>
<td>GPRD-Cox</td>
<td>0.75 (0.65 to 0.86)*</td>
</tr>
<tr>
<td>GPRD-PERR</td>
<td></td>
</tr>
<tr>
<td>GPRD-no missing‡</td>
<td>0.91 (0.79 to 1.05)</td>
</tr>
<tr>
<td>WH-hysterectomy</td>
<td></td>
</tr>
<tr>
<td>Trial</td>
<td>1.01 (0.88 to 1.22)</td>
</tr>
<tr>
<td>GPRD-Cox</td>
<td>0.68 (0.57 to 0.81)*</td>
</tr>
<tr>
<td>GPRD-PERR</td>
<td></td>
</tr>
<tr>
<td>GPRD-no missing‡</td>
<td>0.82 (0.66 to 1.02)</td>
</tr>
<tr>
<td>A5</td>
<td></td>
</tr>
<tr>
<td>Trial</td>
<td>0.70 (0.58 to 0.85)</td>
</tr>
<tr>
<td>GPRD-Cox</td>
<td>0.71 (0.53 to 0.96)</td>
</tr>
<tr>
<td>GPRD-PERR</td>
<td></td>
</tr>
<tr>
<td>HOPE</td>
<td></td>
</tr>
<tr>
<td>Trial</td>
<td>0.84 (0.75 to 0.95)</td>
</tr>
<tr>
<td>GPRD-Cox</td>
<td>0.94 (0.85 to 1.03)</td>
</tr>
<tr>
<td>GPRD-PERR</td>
<td></td>
</tr>
<tr>
<td>EUROPA</td>
<td></td>
</tr>
<tr>
<td>Trial</td>
<td>0.89 (0.77 to 1.02)</td>
</tr>
<tr>
<td>GPRD-Cox</td>
<td>1.06 (0.95 to 1.19)</td>
</tr>
<tr>
<td>GPRD-PERR</td>
<td></td>
</tr>
</tbody>
</table>

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. 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. Recently, propensity scores combined with regression calibration were used to address unobserved variables under certain conditions.

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. Its applicability and validity for studies of therapeutic efficacy have not been widely examined. Some have suggested that this technique is most suited to address health policy issues rather than specific clinical issues of treatment effectiveness.

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. 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.

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.

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.

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.

17 Concato J, Shah N, Horwitz RI. Randomized, controlled trials.
16 McKee M, Britton A, Black N, McPherson K, Sanderson C, Bain C.
19 Grodstein F, Manson JE, Colditz GA, Willett WC, Speizer FE,
26 Tannen RL, Weiner MG, Xie D, Barnhart K. A simulation using data from
12 McMahon AD. Observation and experiment with the efficacy of drugs:
13 MacMahon S, Collins R. Reliable assessment of the effects of
29 Tannen RL, Weiner MG, Xie D, Barnhart K. Estrogen affects post-
21 Vandenbroucke JP. When are observational studies as credible as
28 Weiner MG, Xie D, Tannen RL. Replication of the Scandinavian
25 Tannen RL, Weiner MG, Marcus SM. Simulation of the Syst-Eur
22 Coronary artery surgery study (CASS): a randomized trial of coronary
27 Tannen RL, Weiner MG, Xie D, Barnhart K. Estrogen affects post-
1999;319:312-5.
2000;19:17-34.
2007;297:278-95.

58 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.
59 Accepted: 19 October 2008