Received: 1 December 2015 | Revised: 15 September 2016 | Accepted: 22 September 2016

DOI 10.1002/lrh2.10016

Learning Health Systems

RESEARCH REPORT

A new method to address unmeasured confounding of mortality in observational studies

Richard Tannen[1] | **Menggang Yu**[2]

[1] Clinical Epidemiology, University of Pennsylvania, Philadelphia, Pennsylvania

[2] Department of Biostatistics and Informatics, University of Wisconsin Madison, Madison, Wisconsin

Correspondence
Richard Tannen, University of Pennsylvania, Clinical Epidemiology, 423 Guardian Dr, Philadelphia, PA 19104.
Email: tannen@mail.med.upenn.edu

Abstract

Introduction The prior event rate ratio (PERR) overcomes "unmeasured confounding" by adjusting study outcomes for all confounding (measured and unmeasured) by comparing exposed to unexposed cohort outcomes prior to study entry when neither group is receiving treatment. However, PERR cannot address "unmeasured confounding" of death since prior events cannot occur.

Methods This study's goal is to determine whether a new method, built on the concepts that led to the PERR development, reliably overcomes unmeasured confounding for death. In contrast to the PERR, which precedes study onset, the new mortality analysis uses exposed and unexposed cohorts, not taking the treatment medication, at the end of the study. It is called the post-treated event rate ratio (PTERR).

Results Theoretical and simulation studies were used to evaluate the likelihood for reliable results using of this new analytic strategy. Also, prior empiric studies, which used both the UK GPRD and THIN databases to examine and validate the PERR method, were used to ascertain the validity of the PTERR method.

Conclusion In the aggregate the results provide strong evidence that the PTERR method to evaluate unmeasured confounding will be a valuable analytic tool.

KEYWORDS

medical record database, PERR, unmeasured confounding

1 | INTRODUCTION

The prior event rate ratio (PERR) method developed by our group may effectively cope with "unmeasured confounding" by adjusting study outcomes for all confounding (both measured and unmeasured) by comparing exposed to unexposed cohort outcomes prior to study entry when neither group is receiving treatment.[1–9] Discovery of the PERR method resulted from our initial studies designed to determine whether data obtained from a large electronic medical record database could produce valid results for outcomes research. To address this issue exposed and unexposed cohorts extracted from the database were designed to replicate those from previously performed randomized controlled studies, and the results from the database studies were compared to the RCT studies to determine validity. In some instances the database and RCT results were similar, but in others they were significantly different, suggesting the presence of "unrecognized confounding". This led to identification of the PERR method, which overcame unrecognized confounding and produced similar database and RCT outcome results. In addition to evaluation by empiric studies, subsequent theoretical derivations and simulations all support the usefulness of this analytic technique.

However, PERR cannot address "unmeasured confounding" of death since prior events cannot occur. Since mortality is an important outcome in many studies, we were motivated to determine whether a strategy could be developed to overcome "unmeasured confounding" for this outcome. Our recent publication, which found that pioglitazone and rosiglitazone significantly reduced mortality in an unselected cohort of subjects with type 2 diabetes mellitus[8] provided

an excellent opportunity to determine whether a new method could overcome or mitigate the possible effect of "unmeasured confounding" for mortality. This study was particularly suitable for exploration, since the cohorts were large and decreased mortality results, adjusted using standard statistical methods, were of substantial magnitude. Mortality was discussed cautiously in the publication, since these results could be spurious because of "unmeasured confounding."

The aim of this current study is to examine the feasibility of a new method to overcome "unmeasured confounding" for the mortality outcome. The new method to analyze mortality, called the post-treated event rate ratio (PTERR), builds on the concepts that led to the development of the PERR method. However, PTERR adjusts comparison between exposed and unexposed cohorts during the period following the exposure to treatment. This manuscript addresses empiric, theoretical, and simulation studies evaluating this new analytic strategy and considers its advantages and shortcomings.

2 | METHODS

The methods in this study are similar to those used for determining the validity of the PERR method. Empiric studies and theoretical and simulation studies have been performed in a fashion similar to the PERR method. [1–9]

2.1 | Empiric studies

In addition to the pioglitazone/rosiglitazone study, PTERR will be evaluated empirically using other prior studies, which were the basis for development of the PERR method. The thiazolidenedione studies, performed comparably to our other RCT replication studies, used The UK Health Improvement Network database, which is similar to UK General Practice Research Database (GPRD). Prior studies, which used GPRD, included replications of the HOPE, EUROPA, WHI intact uterus, and the WHI hysterectomy RCT's.[1–3,7] The 4S study was not used because the cohort was too small.[6]

All the GPRD or The UK Health Improvement Network database studies replicated previously performed RCT's to the extent feasible except for randomization. Exposed and unexposed cohorts were selected to mimic entry requirements of the RCT. The exposed cohort included all database subjects meeting the entry criteria and began taking the study medication within a predefined recruitment time interval. Study start time was the date when therapy began. Potential unexposed cohort included all subjects that met RCT entry criteria but did not take study medication during the recruitment interval. The final unexposed cohort was selected by computer-based random matching for age and sex to each exposed subject, and their start time defined as identical to the exposed matched subject. Both cohorts had a predefined similar study end time. Subjects that left the database or died prior to study end date had observations terminated at that time. Thus, study time frames were similar for exposed and unexposed subjects.

Database studies were analyzed using both a simulated "intention-to-treat" analysis, where subjects were analyzed until a predefined study end date or death, and also a simulated "as-treated" analysis where a subject's study ended if exposed stopped or unexposed started medication prior to the predefined end date.

One advantage of our large database studies was the ability to capture information on rate of an event both for the as-treated and post-treated study periods. The exposed subjects began their post-treated study period when they stopped taking the drug. Unexposed subjects were randomly matched to the exposed subjects on the basis of age and sex, and their as-treated period ended if they began taking study medication (an infrequent occurrence). The intention-to-treat design ended at a similar predesigned stop point for both the exposed and unexposed cohorts or at the time of death, if it preceded the defined study end date.

2.2 | Death analysis strategy

The death analytic strategy takes advantage of both the intention-to-treat and as-treated analyses. In every study a substantial portion of the exposed cohort had a longer study interval for the intention-to-treat than the as-treated paradigm. Thus, an interval of study at the end, where the exposed cohort no longer received treatment, could be defined. Absence of medication use by the unexposed cohort was essentially similar between the intention-to-treat and as-treated analyses.

The following strategy was used (Figure 1). All exposed subjects were included in the as-treated period, whereas their post-treated period included only subjects whose intention-to-treat duration exceeded their as-treated duration. Since the unexposed subjects in our original analyses had a much longer as-treated duration than the exposed subjects, this duration required adjustment to similar durations to perform the death analysis. Therefore, the unexposed subjects compared to exposed subjects during the as-treated period were limited to a time duration comparable to the exposed subjects (adjusted as-treated duration), done by limiting duration of the unexposed subjects to the end date of their matched exposed subject as-treated interval. Their post-treated period encompassed the interval between their entire as-treated duration and their adjusted as-treated duration. Furthermore, unexposed post-treated subjects were restricted to those that matched to the exposed post-treated subjects. This produced reasonably similar dates for start and end times for post-treated exposed and unexposed subjects.

2.3 | Analyses

Analytics for evaluation of the Death outcome were as follows:

- Cox univariable (using only the exposure) and multivariable (using both exposure and known confounders) hazard ratios (HRs) for the as-treated period and univariable HR's for the post-treated period.
- The PTERR adjustment: defined as a ratio of Cox univariable HR from the as-treated period by the Cox univariable HR from the post-treated period.
- Results of the Cox multivariable analysis and the PTERR method also were compared.

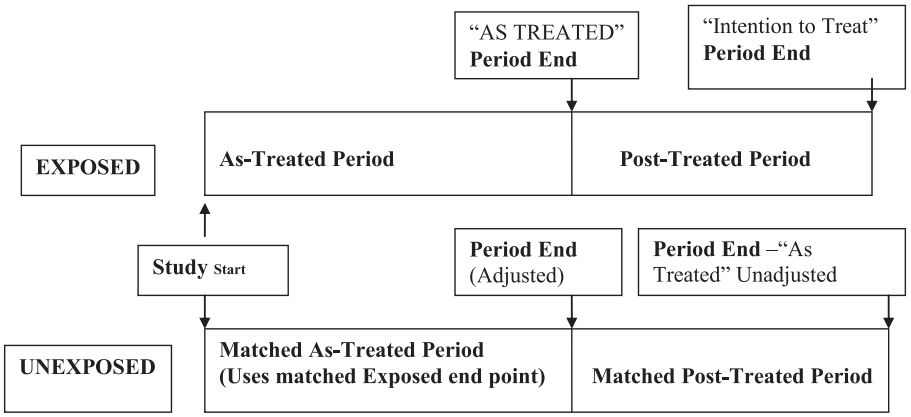

FIGURE 1 Death strategy figure. The empiric studies were performed in the following fashion: both the exposed and unexposed cohorts were analyzed during an "as-treated" period and a "post-treated" period. The exposed "as-treated" period included all subjects until the defined end of the as-treated period, which occurred when they discontinued the study medication. The post-treated period ended when the "intention to treat" analysis ended. Thus, subjects in the post-treated period were those that no longer received the study medication. The unexposed "as-treated" period was designed to approximate the duration of the exposed "as-treated" period. Thus, the "as-treated" adjusted duration of the unexposed was limited to the duration of their matched exposed subject's "as treated" duration. The unexposed post-treated period began after the adjusted "as-treated" period and ceased at the end of their entire "as-treated" duration. Another requirement for inclusion in the unexposed post-treated period was that their "matched" exposed subject was included in the exposed cohort post-treated period. This assured reasonably similar post-treated start and end times for both the unexposed and exposed cohorts

- PTERR results also were compared to the RCT mortality HRs, or to observational study HRs if RCT data were unavailable.

Incidence rates (IRs) were calculated for the as-treated and post-treated periods. Incidence rate ratios (IRRs) were analyzed for both the exposed and unexposed cohorts, comparing the post-treated to the as-treated period for each individual cohort.

Baseline confounders were compared between the as-treated and post-treated periods.

3 | THEORETICAL DERIVATIONS AND SIMULATIONS OF THE PTERR METHOD

To accommodate different confounding effects, we assume the corresponding models for different periods in Table 1. In this setup, the exposure effect β is not estimable directly from the exposed group because of the confounder U_2. In particular, when the mean of U_2 is nonzero, directly applying the Cox model only to the exposed group will lead to a biased estimator of β. The PTERR method alleviates this confounding issue by using a constructed unexposed cohort and applying a difference-in-difference technique. Note that the unmeasured confounders V_1 and V_2 for the unexposed may differ from U_1 and U_2 for the exposed. The estimate for the exposure effect $\widehat{\beta}_{As-Trt}$ on the basis of the as-treated period using the Cox univariable model contains both the true exposure effect β and the different confounding effects from U_1 and V_1. This difference is adjusted by further fitting a Cox univariable model for the post-treated period, which leads to another hazard estimate $\widehat{\beta}_{Post-Trt}$ between the exposed and unexposed cohorts. The PSERR estimate for β is $\widehat{\beta}_{PTERR} = \widehat{\beta}_{As-Trt} - \widehat{\beta}_{Post-Trt}$.

TABLE 1 Underlying hazard models for PTERR method investigation

	As-treated Interval	Post-treated Interval
Exposed	$\lambda_0(t)e^{U1+\beta}$	$\lambda_0(t)e^{U1+U2}$
Unexposed	$\lambda_0(t)e^{V1}$	$\lambda_0(t)e^{V1+V2}$

From a general theoretical derivation,[9] when the mortality rate is rare, $\widehat{\beta}_{As-Trt}$ and $\widehat{\beta}_{Post-Trt}$ can be approximated by

$$\widehat{\beta}_{As-Trt} \approx \beta + \log\left\{E\left[e^{U_1-V_1} \middle| T_{E,As-Trt} \geq T_{UnE,As-Trt}, C_E \geq T_{UnE,As-Trt}, C_{UnE} \geq T_{UnE,As-Trt}\right]\right\}$$

where $T_{UnE,\,As-Trt}$ and $T_{E,\,As-Trt}$ represent event variables from the unexposed and exposed cohorts for the as-treated period and C_{UnE} and C_E are the censoring variables correspondingly. The second part of the above equation represents the different confounding effects from U_1 and V_1. Similarly,

$$\widehat{\beta}_{Post-Trt} \approx \log\left\{E\left[e^{U_1-V_1+U_2-V_2} \middle| T_{E,Post-Trt} \geq T_{UnE,Post-Trt}, C_E \geq T_{UnE,Post-Trt}, C_{UnE} \geq T_{UnE,Post-Trt}\right]\right\}$$

As a result, the proposed estimate $\widehat{\beta}_{As-Trt} - \widehat{\beta}_{Post-Trt}$ is nearly unbiased for β when $U_2 - V_2$ is close to zero or has relatively negligible effect when compared with $U_1 - V_1$ or β.

4 | RESULTS

4.1 | Empiric studies

Prior empiric database studies were used to determine whether PTERR was a valid strategy for dealing with "unmeasured confounding" for mortality analysis. Because of their unique characteristics, study results are described as 3 distinct clusters, before discussing a

combined analysis. The clusters are as follows: thiazolidinedione studies, angiotensin converting enzyme inhibitor (ACEI) studies (HOPE & EUROPA), and women's health initiative (WHI) studies (WHI intact uterus and WHI hysterectomy).

4.2 | Thiazolidenedione studies

Death results from these studies stimulated development of a method to overcome the effect of "unmeasured confounding" on the mortality analysis, which represents an expansion of our initial work that led to development of the PERR method.[1–9] The pioglitazone replication study used the PROACTIVE RCT as the comparator for the database replication study.[10] Rosiglitazone replication study was performed in exactly the same fashion to compare results between the 2 thiazolidenedione medications.

Expanded studies of both medications also were performed, which used the same age criteria for entry as the replication studies, but had no other entry or exclusion criteria. The expanded studies had large cohorts and both exhibited substantial reductions in mortality. No specific RCT comparisons were available for the thiazolidenedione expanded database studies. However, the RECORD RCT, which used rosiglitazone medication, had entry criteria and baseline data reasonably similar to the rosiglitazone expanded study.[11]

Table 2 shows IR comparing as-treated to post-treated periods for exposed and unexposed cohorts. Expanded rosiglitazone and pioglitazone studies exhibited similar results. The exposed cohort IR was significantly lower in the as-treated than post-treated period, whereas the unexposed cohort results were virtually identical for the 2 periods.

Table 3 shows comparisons of the HRs between the exposed and unexposed cohorts. The as-treated period, Cox univariable (unadjusted), Cox multivariable (adjusted), and PTERR HR's are shown. In both the rosiglitazone and pioglitazone studies all the as-treated period HR's were decreased significantly. No difference existed between the Cox multivariable and PTERR results, consistent with absence of "unrecognized confounding."

Rosiglitazone and pioglitazone replication studies were largely similar to the expanded studies (Table 2). In both studies the IR's for the exposed cohorts were significantly lower in the as-treated than post-treated periods. There were no significant differences in the

TABLE 2 Mortality: IR comparison between as RX and post-study periods

	AS RX[a] Period				Post-treated Period				AS RX[a] vs Post-treated Period		
Study	Subject, N	Duration 100 pt yr[b]	Death, N	IR	Subject, N	Duration 100 pt yr[b]	Death, N	IR	IRR	95% CI	*P*
TZD											
ROS expanded											
Exposed	10,524	282.0	400	1.42	4,913	129.8	355	2.73	0.52	0.45,0.60	<.01
Unexposed	31,716	736.7	2049	2.78	12,774	363.7	1,013	2.79	0.99	0.93,1.08	.97
PIO expanded											
Exposed	3,844	97.6	120	1.23	1,808	48.6	135	2.77	0.44	0.34,0.57	<.01
Unexposed	10,994	241.9	668	2.76	4,645	137.5	384	2.79	0.99	0.87,1.12	.86
ROS replication											
Exposed	1,935	50.5	115	2.28	951	23.4	112	4.79	0.48	0.37,0.62	.01
Unexposed	4,951	110.8	398	3.59	2,046	52.7	180	3.42	1.05	0.88,1.25	.58
PIO replication											
Exposed	708	18.0	33	1.83	329	8.62	39	4.52	0.41	0.25,0.64	<.01
Unexposed	1611	36.7	109	2.96	630	17.7	69	3.90	0.76	0.56,1.03	.08
ACEI											
HOPE											
Exposed	9,235	249.3	1345	5.40	2,638	67.5	434	6.42	0.84	0.75,0.93	<.01
Unexposed	15,717	331.6	1379	4.15	3,656	99.4	445	4.48	0.93	0.83,1.03	.18
EUROPA											
Exposed	7,253	203.1	898	4.42	2,347	60.3	324	5.37	0.82	0.72,0.93	<.01
Unexposed	12,102	265.0	900	3.40	3,234	89.2	315	3.53	0.96	0.84,1.09	.55
WHI											
WHI 50-70 y											
Exposed	34,006	1292.2	243	0.19	13,813	448.9	248	0.55	0.34	0.29,0.41	<.01
Unexposed	64,226	1746.9	832	0.48	19,784	761.7	428	0.56	0.85	0.75,0.95	<.01
HYST 50-70 y											
Exposed	13,369	576.2	115	0.20	4,886	164.2	106	0.65	0.31	0.24,0.40	<.01
Unexposed	20,206	594.1	289	0.49	5,144	211.4	134	0.63	0.77	0.63,0.94	.011

[a]As RX means "as treated."

[b]100 pt yr means "100 patient years."

Abbreviations: IR, incidence rate; IRR, incidence rate ratio. TZD, thiazolidenedione.

TABLE 3 Mortality: comparison between exposed and unexposed cohorts

Database Study	As-treated Period Hazard Ratios (HR): Exposed/Unexposed				Post-treated Period HR Cox univariable (PTERR)	Comparison Study		
	Cox Univariable. (95% CI)	Cox Multivariable (95% CI)	$PTERR_a$ adj (95% CI)	Cox Multivariable vs PTERR adj P-value		Study	HR (95% CI)	vs PTERR adj P-value
TZD								
ROS expanded	0.50 (0.45,0.56)	0.56 (0.50-0.63)	0.53 (0.45,0.62)	NS	0.95 (0.84,1.07)	RECORD RCT	0.87 (0.70,1.08)	<0.01
PIO expanded	0.44 (0.36,0.53)	0.47 (0.38,0.58)	0.44 (0.34,0.58)	NS	0.99 (0.82,1.21)			
ROS replication	0.61 (0.50,0.75) (0.50,0.75)	0.67 (0.53-0.84)	0.44 (0.32,0.58)	0.026	1.39 (1.10,1.76)			
PIO replication	0.62 (0.42,0.92)	0.61 (0.40,0.94)	0.54 (0.32,0.88)	NS	1.16 (0.78,1.71)	PROACTIVE RCT	0.96 (0.78,1.10)	0.011
ACEI								
HOPE	1.26 (1.17,1.36)	1.13 (1.05,1.26)	0.88 (0.75,1.02)	<0.01	1.43 (1.26,1.64)	HOPE RCT	0.84 (0.75,0.95)	NS
EUROPA	1.3 (1.18,1.42)	1.1 (1.01,1.22)	0.85 (0.72,1.02)	<0.01	1.52 (1.30,1.78)	EUROPA RCT	0.89 (0.77,1.02)	NS
WHI								
WHI 50-70 y	0.38 (0.33,0.44)	0.55 (0.46,0.62)	0.38 (0.31,0.48)	<0.01	0.99 (0.85,1.16)	WHI 50-70	0.98 (0.77,1.25)	<0.01
WHI 50-60 y	0.42 (0.36-0.49)	0.55 (0.47-0.66)	0.39 (0.29,0.50)	0.027	1.08 0.89,1.31	WHI 50-60	0.69 (0.44,1.07)	NS
HYST 50-70 y	0.37 (0.30,0.46)	0.44 (0.35,0.54)	0.39 (0.28,0.54)	NS	1.02 (0.79,1.31)	WHI HYST 50-70	0.94 (0.75,1.16)	<0.01
HYST 50-60 yrs.	0.38 (0.29,0.50	0.43 (0.26,0.65)	0.44 (0.28,0.70	NS	0.86 (0.61,1.20)	WHI HYST 50-60	0.71 (0.46,1.11)	NS

Abbreviation: PTERR, post-treated event rate ratio.

unexposed cohort between the as-treated and post-treated periods, but the IRR in the pioglitazone study was lower than 1.0 in contrast to another studies. The much lower pioglitazone cohort size than another studies may account for this difference.

All HR's for the as-treated period (Table 3) were significantly decreased in both the rosiglitazone and pioglitazone replication studies; however, the PTERR adjusted HR in the rosiglitazone study was significantly lower than the Cox multivariable HR, suggesting the possibility of "unmeasured confounding."

Comparison of the pioglitazone replication study with the PROACTIVE RCT suggests that the database results differ. The PROACTIVE RCT exhibited a slight decrease in mortality, which was not significant.

4.3 | Angiotensin converting enzyme inhibitor studies (HOPE and EUROPA RCT's)

The original HOPE and EUROPA database studies both compared the Exposed group that took any dose of ACEI, with a subset that took greater than 4.0 mg/day ramipril equivalent.[7,11–13] Studies with either dosage had similar results. The mortality analysis uses the much larger cohort that took any dose of ACEI. An additional advantage was that matching the unexposed to exposed was complete with the any dose cohort.

The HOPE and EUROPA mortality study results were very similar. In both studies the exposed cohorts exhibited a significant decrease in IR during the as-treated as compared with the post-treated period (Table 2—IRR), whereas the unexposed cohort's IRs were similar in the 2 periods.

In both studies (Table 3) the PTERR HR's for the as-treated period were significantly decreased in comparison with the Cox multivariable HR. Furthermore, the PTERR HR's were virtually identical to the Hope and Europa RCT results (Table 3). This provides strong support for the presence of "unrecognized confounding" in both these studies.

Of interest both the HOPE and EUROPA studies also demonstrated "unrecognized confounding" for another outcomes (myocardial infarction, stroke, and coronary revascularization [coronary artery bypass operation or percutaneous transluminal coronary angioplasty]), which were corrected by applying the PERR adjustment method.[1] In both studies prior events in the exposed were greater than those in the unexposed producing a decrease in the PERR adjusted outcomes. This is similar to the mortality analysis where death in the post-treated period was higher in the exposed than unexposed (Table 2), resulting in a decrease in the PTERR adjusted mortality compared with Cox adjusted mortality (Table 3).

4.4 | WHI intact uterus and hysterctomy RCT'S

The WHI intact uterus and hysterectomy RCT's both had an age distribution dictated by study design that did not reflect the ages of women in the nonselected population treated with post-menopausal hormone therapy.[14–18] Both WHI RCT's were heavily weighted to older women. Therefore, our database replications of the WHI exhibited important differences in age of both exposed and unexposed subjects than the RCT's.[2–4]

Recognizing the age problem, the WHI investigators performed post-hoc analyses of different age groups using the original WHI data. These post-hoc analyses lose the advantage of randomization and should be considered observational studies.[18]

The RCT post-hoc analyses suggest that younger women have a reduction in death in response to post-menopausal therapy. The results for the 50- to 60-year-old cohorts in both the intact uterus and hysterectomy studies exhibit an HR of 0.69 (0.44, 1.07) and 0.71 (0.46, 1.11), respectively, neither achieving statistical significance. Combining these 2 studies produces a statistically significant HR of 0.70 (0.51, 0.96). By contrast the 50- to 70-year-old subset has HR's of almost 1.0.[18]

To more closely, but not completely, replicate these findings, our database studies were analyzed using women 50 to 70 years of age and also 50 to 60 years of age. These cohorts were developed by combining the 50- to 55-year-old and 55- to 75-year-old cohorts from our published studies, which were derived together.[2–5]

The intact uterus 50- to 70-year-old exposed cohort had a lower IR during the as-treated than post-treated period (Table 2). The unexposed cohort also had a significantly lower IR during the as-treated than post-treated period (Table 2), but the IR during the as-treated period was much higher than the exposed cohort. The hysterectomy study results were very similar to the intact uterus results.

The HR's during the as-treated period (Table 3) from the Intact Uterus database study all were significantly decreased, but were lower with PTERR than with the Cox multivariable (adjusted) analysis. Hysterectomy study results did not exhibit significant differences between the Cox and PTERR adjusted results.

Consistent with the post-hoc analysis of the WHI RCT 50- to 60-year studies, our results reveal a significant reduction in mortality for both the WHI intact uterus and WHI hysterectomy studies, and also show reductions in the 50- to 70-year cohorts. The reduction is similar for both Cox multivariable as-treated period and the PTERR analysis in the hysterectomy study, suggesting no unrecognized confounding; however, the PTERR is lower than the Cox multivariable results in the intact uterus analyses. The death reduction is larger than the WHI post-hoc analyses, but average age in our study is lower than the WHI RCT post-hoc analysis.

In contrast to the PERR method, which uses data prior to study entry, the PTERR method that uses data after exposure to medication could have the exposed cohort post-treated period impacted by a residual effect of prior medication ingestion. This issue was addressed by assessing several aspects of our results as follows:

1. If the unexposed cohort exhibits reasonably similar IR's when the as-treated period is compared to the post-treated period, this assures a stable IR over the duration of the study. This occurred in 6 of the 8 cohorts studied (Table 2).
2. Absence of a significant difference between the IR for the exposed compared to unexposed during the post-treated period provides strong evidence against a residual medication effect in the exposed cohort, which occurred in 5 of 8 cohorts. (Table 2)
3. When a significant difference exists between the exposed and unexposed IR during the post-treated period a residual medication effect would require a similar directional IR change during the as-treated period. This only occurred in the HOPE and EUROPA studies, where the Cox univariable HR was increased during both the post-treated and the as-treated period (Table 2). This potentially could be due to a residual medication effect; however, since the increased HR in the post-treated period was much higher than during the as-treated period, it is unlikely that the residual medication effect would be greater than the effect while medication was being taken. An alternative explanation would be the presence of "unmeasured confounding," which indeed was shown to account for this effect in both the HOPE and EUROPA studies

Therefore, there does not appear to be any evidence of a residual medication effect in any of the studies performed. In the aggregate, our studies provide strong evidence that the PTERR method to evaluate unmeasured confounding will be a valuable analytic tool.

4.5 | Baseline confounders

To ascertain that the cohort characteristics were similar in the as-treated and post-treated periods, baseline confounders were compared (including prior medication, prior health events, and demographics) between as-treated and post-treated periods for both the exposed and unexposed of all study cohorts. In the aggregate there were reasonable similarities between the 2 periods for both the exposed and unexposed cohorts.

4.6 | Simulation results of the PTERR method

To investigate how well or sensitive these estimates are to the rare event assumption, we conducted simulation studies by varying degrees of the rare event rates. We generated data according to Table 1 where the baseline hazard $\lambda_0(t)$ is taken as e^{α_0} with different values of α_0 representing different baseline rates of mortality. The unmeasured confounders U_1,U_2, V_1, and V_2 are generated from normal distributions with standard deviation of 0.5. The means of U_1 and V_1 are fixed at 0 and those of U_2 and V_2 changed from different values for evaluation of PTERR performance. The true treatment effect β is taken as −1. We simulated 5000 data sets with sample size of 5000. The PTERR was performed, and the estimates were summarized in Figure 2. From the figure, we can see that PTERR works well. In this particular setting, as the means of U_2 and V_2, denoted as μ_2 and ν_2, shift from −0.4 to 0.4, the bias remained small with its magnitude less than 0.03. The bias also seems to increase with larger values of μ_2 and ν_2, and with larger values of α_0. We have investigated more settings and the main message seems to hold from our investigation.

5 | DISCUSSION

The empiric, theoretical derivations and simulations of the PTERR method to cope with "unrecognized confounding" for the death outcome support the conclusion that it is a valuable new analytic tool.

Following our theoretical work for the PERR method, we derived large sample properties of the PTERR method.[9] The properties demonstrate that PTERR can correct bias because of unmeasured

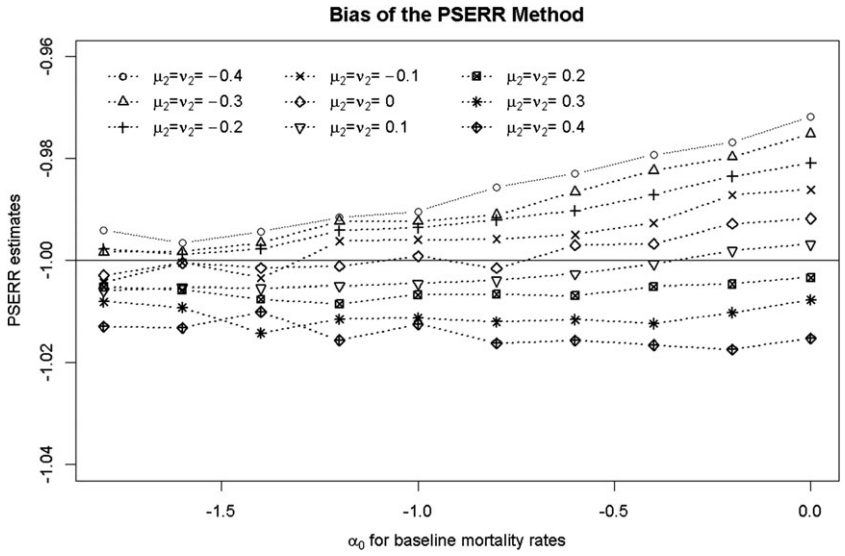

FIGURE 2 The PSERR simulation study figure. The baseline hazard is taken as $e^{\alpha o}$ where different values of α_0 were used to represent different baseline rates of mortality and listed in the x-axis. The unmeasured confounders U_1,U_2, V_1, and V_2 are generated from normal distributions with standard deviation of 0.5. The means of U_1 and V_1 are fixed at 0 and those of U_2 and V_2 denoted as μ_2 and v_2, shift from −0.4 to 0.4. The true treatment effect β is taken as −1. The average PSERR estimates from 5000 simulated data sets are plotted. The bias is small and seems to increase with larger values of μ_2 and v_2, and with larger values of α_0

confounding when the influences from unmeasured confounders on the as-treated and the post-treated periods are relatively similar between the exposed and unexposed groups. This can happen when the unmeasured confounders affect the hazards are temporally stable between the as-treated and the post-treated periods, even though they can be different between the 2 groups, or it can happen when confounding effects change temporally, but the changes are relatively stable between the 2 groups. Considerable bias can result in other settings where the unmeasured confounding effects between the as-treated and the post-treated periods differ greatly between the 2 groups as compared with the true exposure effect. Similar conclusions were also made in the PERR method.[9]

Ideally, similar to our previous empiric studies of the PERR method, comparison with a replicated RCT would have been used to confirm validity.[1] This was not feasible except for the HOPE and EUROPA studies, which both demonstrated "unmeasured confounding" for death that was corrected with the PTERR adjustment to values virtually identical to the replicated RCT.[7] Data from observational studies were required to evaluate reliability of the death outcome for another database studies.

Thiazolidenedione studies used the PROACTIVE RCT to compare with the pioglitazone replication study, but this was the smallest cohort of all the empiric studies, and PROACTIVE RCT results differ from many reported observational studies. The RECORD RCT, which investigated rosiglitazone treatment in subjects with type 2 diabetes mellitus, had entry criteria reasonably similar to the ROS expanded database study. The RECORD entered subjects 40 to 75 years old, 321 in the treated and 323 into the control arm. The ROS expanded study included subjects ages 35 to 75 years, with 30 times greater treated and 100 times greater untreated than record. Mortality reduction in record was reduced but not significantly 0.86 (0.68-1.08); compared to 0.57 (0.50-0.67) in the ROS expanded study. Thus, the trend was similar and lack of significance in RECORD might be explained by the much lower number of subjects.

Several observational studies of thiazolidenediones, reviewed in a recent FDA analysis, suggest a decrease in mortality.[19] Two of the most recent studies addressed subjects with an age range similar to ours. Tzoulaki et al using the GPRD found a decrease with either rosigilitazone 0.88 (0.70-1.09) or pioglitazone 0.69 (0.49-0.98).[20] Habib et al, using the Henry Ford Health System database reported a decrease with rosiglitazone 0.87 (0.54-1.39) and a larger decrease with pioglitazone 0.63 (0.45-0.87).[21] Loebstein et al using the Macabi Health Care Service data investigated rosiglitazone and found no significant change 1.15 (0.85-1.56).[22] Other studies restricted to older subjects (>65 years) report increased mortality, but most are not significant statistically.[23,24] In contrast to our studies, none of these studies provide a rigorous comparison between exposed and unexposed cohorts with similar characteristics in the fashion we have used.

Thus, validity of the thiazolidenedione (TZD) studies cannot be proven rigorously with our studies, but some observational studies are consistent with our results. Another strength of our study is the consistency of the findings with 4 different cohorts of TZD treatment.

The WHI intact uterus and WHI hysterectomy RCT's cannot be used for a rigorous comparison with the method, because the RCT's had a forced age distribution for subjects that does not reflect the actual age distribution of women in the population taking post-menopausal hormone therapy.[14] Thus, subject age distribution in our database studies did not mimic the RCT. However, a post-hoc analysis of the RCT results suggest that younger women in both the WHI intact uterus and hysterctomy RCT's exhibit a reduction in mortality.[18] Although these post-hoc results are not statistically significant, they are similar for both the WHI and WHI hysterectomy cohorts, and combined analysis of the 2 studies achieved statistical significance.[18] In addition, a meta-analysis suggests that women less than 60 years old treated with post-menopausal hormone therapy have a decrease in mortality.[25] Our database studies of both the WHI intact uterus and hysterectomy studies demonstrate a decrease in mortality consistent with WHI post-hoc analyses. Thus, although our WHI database studies cannot be rigorously verified by comparison with an RCT, the observational studies derived from the WHI RCT's are consistent with our results.

The likelihood that "unmeasured confounding" is present is determined by comparing the results of the Cox multivariable HR during the as-treated period with the results of the PTERR adjusted HR. When a

significant difference is found, the presence of "unmeasured confounding" is presumed. This follows a paradigm similar to detecting "unmeasured confounding" for other outcomes using the PERR method. This occurred only in the HOPE and the EUROPA studies (Table 3). Of interest "unmeasured confounding" was also detected for other outcomes in both these studies using the PERR method.

Ideas similar to PERR have appeared in literature as we pointed out in our previous publications.[1,9] In particular, the difference-in-difference (DID) method used in economics has some similarities.[26] However, application of DID in clinical epidemiology is very limited. A series of methods related to "case" cross-over designs[27–29] also employed a somewhat similar idea; however, these designs do not specifically incorporate a control group to capture extra unmeasured confounding. Conceptually, our idea falls under the "negative controls" framework proposed by Lipsitch et al[30] and therefore has the ability to improve causal inference.

As discussed in our prior publications, 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.[1,9]

As confirmed in our prior empiric 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.[31–33]

Instrumental variable analysis and propensity scores combined with regression calibration also have been used to address unmeasured confounding.[34–36]

An instrumental variable analysis requires identification of an appropriate instrument that affects the assignment to treatment but has no direct effect on the outcomes,[36–38] whereas propensity score calibration technique requires the presence of a validation study. 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.

Furthermore, in contrast to the PERR method, the validity of these methods has not been rigorously ascertained.

Although the primary focus of this work was to examine the utility of the PTERR adjustment for analysis of death, some of the specific findings also are of considerable interest. Regarding the thiazolidenedione studies, they represent some of the strongest evidence that both pioglitazone and rosiglitazone reduce mortality. Furthermore, the WHI studies provide strong evidence that younger women respond to post-menopausal therapy with a reduction in mortality.

REFERENCES

1. Tannen RL, Weiner MG, Xie D. Use of primary care electronic medical record database in drug efficacy research on cardiovascular outcomes: comparison of database and randomized controlled trial findings. *BMJ*. 2009;338:b81.
2. Tannen RL, Weiner MG, Xie D, Barnhart K. Simulation of the women's health initiative trial using data from a primary care practice database. *J Clin Epidemiol*. 2007;60:686–695.
3. Tannen RL, Weiner MG, Xie D, Barnhart K. Estrogen affects post-menopausal women differently than estrogen plus progestin therapy. *Hum Reprod*. 2007;22:1769–1777.
4. Weiner MG, Barnhart K, Xie D, Tannen RL. Hormone replacement therapy and coronary heart disease in young women. *Menopause*. 2008;15:86–93.
5. 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;90:258–264.
6. 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;17:661–670.
7. Tannen RL, Weiner MG, Xie D. Replicated studies of two randomized trials of angiotensin converting enzyme inhibitors: further empiric validation of the "prior event rate ratio" to adjust for unmeasured confounding by indication. *Pharmacoepidemiol Drug Saf*. 2008;17:671–685.
8. Tannen RL, Xie D, Wang X, Yu M, Weiner MG. A new "comparative effectiveness" assessment strategy using the THIN database: comparison of the cardiac complications of pioglitazone and rosiglitazone. *Pharmacoepidemiol Drug Saf*. 2013;22:86–97.
9. Yu M, Xie D, Wang X, Weiner MG, Tannen RL. Prior event rate ratio adjustment: numerical studies of a statistical method to address unrecognized confounding in observational studies. *Pharmacoepidemiol Drug Saf*. 2012;21:60–68.
10. Dormandy JA, Charbonnel B, Eckland DJA, et al. Secondary prevention of macrovascular events in patients with type 2 diabetes in the PROactive Study (PROspective pioglitAzone Clinical Trial In macroVascular Events): a randomized controlled trial. *Lancet*. 2005;366:1279–1289.
11. Home PD, Pocock SJ, Beck-Nielsen H, et al. Rosiglitazone evaluated for cardiovascular outcomes in oral agent combination therapy for type 2 diabetes (RECORD): a multicentre, randomised, open-label trial. *Lancet*. 2009;373:2125–2135.
12. The Heart Outcomes Prevention Evaluation Study Investigators. Effects of an angiotensin-converting-enzyme inhibitor, ramipril, on cardiovascular events in high risk patients. *New Engl J Med*. 2000;342:145–153.
13. The 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;362:782–788.
14. The Women's Health Initiative Study Group. Design of the Women's Health Initiative clinical trial and observational study. *Control Clin Trials*. 1998;19:61–109.
15. Writing Group for theWomen'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;288:321–333.
16. The 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;291:1701–1712.
17. Anderson GL, Limacher M, Assaf AR, et al. Effects of conjugated equine estrogen in postmenopausal women with hysterectomy: the Women's Health Initiative randomized controlled trial. *JAMA*. 2004;291:1701–1712.
18. Rossouw JE, Prentice RL, Manson JE, et al. Postmenopausal hormone therapy and risk of cardiovascular disease by age and years since menopause. *JAMA*. 2007;297:1465–1477.
19. U.S. Food and Drug Administration. Division of metabolism and endocrinology productsoffice of drug evaluation and research. Advisory Committee Meeting Avandia® (rosiglitazone maleate), June 5 and 6, 2013.

20. Tzoulaki I, Molokhia M, Curcin V, et al. Risk of cardiovascular disease and all cause mortality among patients with type 2 diabetes prescribed oral antidiabetes drugs: retrospective cohort study using the UK general practice research database. *BMJ*. 2009;339:b4731.
21. Habib ZA, Tzogias L, Havstad SL, et al. Relationship between thiazolidinedione use and cardiovascular outcomes and all-cause mortality among patients with diabetes: a time-updated propensity analysis. *Pharmacoepidemiol Drug Saf*. 2009;18:437–447.
22. Loebstein R, Dushinat M, Vesterman-Landes J, et al. Database evaluation of the effects of long-term rosiglitazone treatment on cardiovascular outcomes in patients with type 2 diabetes. *J Clin Pharmacol*. 2011;51:173–180.
23. Lipscombe LL, Gomes T, Levesque LE, Jux JE, Juurlink DN, Alter DA. Thiazolidinediones and cardiovascular outcomes in older patients with Diabetes. *JAMA*. 2007;298:2634–2643.
24. Winkelmayer WC, Setoguchi S, Levin R, Solomon DH. Comparison of cardiovascular outcomes in elderly patients with diabetes who initiated rosiglitazone vs pioglitazone therapy. *Arch Intern Med*. 2008;168:2368–2375.
25. Salpeter SR, Walsh JME, Greyber E, Ormiston TM, Salpeter EE. Mortality associated with hormone replacement therapy in younger and older women. *A Meta-analysis J Gen Intern Med*. 2004;19:791–804.
26. Athey S, Imbens GW. Identification and inference in nonlinear difference-in-differences models. *Econometrica*. 2006;74:431–497.
27. Navidi W. Bidirectional case-crossover designs for exposures with time trends. *Biometrics*. 1998;54(2):596–605.
28. Suissa S. The case time control design. *Epidemiology*. 1995;6:248–253.
29. Suissa S, Dell'Aniello S, Carlos Martinez C. The multitime case-control design for time-varying exposures. *Epidemiology*. 2010;21:876–883.
30. Lipsitch M, Tchetgen E, Cohen T. A tool for detecting confounding and bias in observational studies. *Epidemiology*. 2010;21:383–388.
31. Rosenbaum PR. Discussing hiddenbias in observational studies. *Ann Intern Med*. 1991;115:901–905.
32. Rubin DB. Estimating causal effects from large data sets using propensity scores. *Ann Intern Med*. 1997;127:757–763.
33. Braitman LE, Rosenbaum PR. Rare outcomes, common treatments: analytic strategies using propensity scores. *Ann Intern Med*. 2002;137:693–695.
34. McMahon AD. Approaches to combat confounding by indication in observational studies of intended drug effects. *Pharmacoepidemiol Drug Saf*. 2003;12:551–558.
35. Sturmer T, Schneeweiss S, Rothman KJ, Avorn J, Glynn RJ. Performance of propensity score calibration. A simulation study. *Am J Epidemiol*. 2007;165:1110–1118.
36. McClellan M, McNeil BJ, Newhouse JP. Doesmore intensive treatment of acutemyocardial infarction intheelderly reducemortality? Analysis using instrumental variables. *JAMA*. 1994;272:859–866.
37. Newhouse JP, McClellan M. Econometrics in outcomes research: the use of instrumental variables. *Annu Rev Public Health*. 1998;19:17–34.
38. Sukel TA, Fisher ES, Wennberg DE, Alter DA, Gottleib DJ, Vermeulen MJ. Analysis of observational studies in the presence of treatment selection bias. Effect of invasive cardiacmanagement onAMI survival using propensity score and instrumental variable methods. *JAMA*. 2007;297:278–285.

How to cite this article: Tannen R, Yu M. A new method to address unmeasured confounding of mortality in observational studies. *Learn Health Sys*. 2017;1:e10016. doi: 10.1002/lrh2.10016