



Practice of Epidemiology

Comparison of Calipers for Matching on the Disease Risk Score

John G. Connolly* and Joshua J. Gagne

* Correspondence to John G. Connolly, Division of Pharmacoepidemiology and Pharmacoeconomics, Brigham and Women's Hospital, 1620 Tremont Street, Suite 3030, Boston, MA 02120 (e-mail: joc121@mail.harvard.edu).

Initially submitted June 6, 2015; accepted for publication October 28, 2015.

Previous studies have compared calipers for propensity score (PS) matching, but none have considered calipers for matching on the disease risk score (DRS). We used Medicare claims data to perform 3 cohort studies of medication initiators: a study of raloxifene versus alendronate in 1-year nonvertebral fracture risk, a study of cyclooxygenase 2 inhibitors versus nonselective nonsteroidal antiinflammatory medications in 6-month gastrointestinal bleeding, and a study of simvastatin + ezetimibe versus simvastatin alone in 6-month cardiovascular outcomes. The study periods for each cohort were 1998 through 2005, 1999 through 2002, and 2004 through 2005, respectively. In each cohort, we calculated 1) a DRS, 2) a prognostic PS which included the DRS as the independent variable in a PS model, and 3) the PS for each patient. We then nearest-neighbor matched on each score in a variable ratio and a fixed ratio within 8 calipers based on the standard deviation of the logit and the natural score scale. When variable ratio matching on the DRS, a caliper of 0.05 on the natural scale performed poorly when the outcome was rare. The prognostic PS did not appear to offer any consistent practical benefits over matching on the DRS directly. In general, logit-based calipers or calipers smaller than 0.05 on the natural scale performed well when DRS matching in all examples.

calipers; cohort studies; confounding (epidemiology); disease risk score; epidemiologic methods; matching; prognostic propensity score; propensity score

Abbreviations: CI, confidence interval; COX-2, cyclooxygenase 2; DRS, disease risk score; ICD-9, *International Classification of Diseases, Ninth Revision*; ns-NSAID, nonselective nonsteroidal antiinflammatory drug; OR, odds ratio; PPS, prognostic propensity score; PS, propensity score.

Propensity scores (PS) are commonly used in database studies of drug effects to adjust for a large number of confounders, particularly when there are few outcome events. One of the most common methods for utilizing the PS, defined as the probability of exposure given a subject's observed characteristics (1), is to match exposed and unexposed subjects with PS values that differ by no more than a prespecified distance, or caliper. Though there is variability in calipers used to match on the PS in the medical literature (2, 3), Rosenbaum and Rubin (4) showed that matching within a caliper of 0.2 times the standard deviation of the logit of the PS would remove 99% of bias due to measured confounding. Calipers defined on the natural PS scale are most commonly used and have been found to perform similarly to logit-based calipers in simulation studies (5).

Disease risk scores (DRS), which combine individual covariates' contributions into a predicted probability of the outcome under a specified reference condition, offer advantages

over PSs in certain situations, such as when the exposure is rare but the outcome is common. Although much early work involving the DRS was in the setting of cohorts of exposed versus unexposed individuals, the DRS has particular advantages when comparing multiple treatment groups (6) or when investigating newly marketed treatments for which there are few exposures (7). Because the DRS is also a balancing score, it can be used like the PS for confounding control (8). While few studies have employed DRS matching, it has the relative advantage over PS matching of being able to include more patients in the analysis because the degree of overlap in the distributions of disease risk between groups is always at least as large as the overlap in PS distributions (9).

Despite the advantages of DRSs and of DRS matching, it is unknown whether recommended calipers for PS matching are appropriate for DRS matching, since the scores exist on different scales. In terms of confounding control, a 1-unit

Table 1. Baseline Characteristics of Initiators of Raloxifene and Alendronate Use (Medicare Claims Data), United States, 1998–2005

Covariate ^a	Raloxifene (n = 9,829)			Alendronate (n = 40,960)		
	Mean (SD)	No.	%	Mean (SD)	No.	%
Demographic factors						
Age, years	77.6 (6.8)			79.1 (6.8)		
Female sex		9,783	99.5		38,319	95.0
White race/ethnicity		9,094	92.5		37,925	92.6
Utilization of health services						
No. of hospitalizations	0.4 (1.0)			0.6 (1.2)		
No. of days hospitalized	3.2 (9.1)			4.7 (11.0)		
No. of different medications used	11.2 (6.0)			11.3 (6.1)		
No. of days in nursing home	1.7 (8.9)			3.5 (13.4)		
No. of physician visits	10.6 (7.2)			10.8 (7.4)		
Bone mineral density testing		3,762	38.3		18,508	45.2
Clinical conditions						
Combined comorbidity score ^b	1.3 (2.2)			1.7 (2.5)		
Alzheimer's disease		631	6.4		3,326	8.1
Cancer		1,962	20.0		8,929	21.8
Cataracts		3,924	39.9		15,321	37.4
COPD		2,165	22.0		10,306	25.2
Crohn's disease		617	6.3		2,270	5.5
Depression		1,135	11.6		4,934	12.1
Diabetes		2,442	24.8		10,931	26.7
Falls		374	3.8		2,332	5.7
Nonvertebral fracture		438	4.5		3,066	7.5
History of fracture		1,294	13.2		8,251	20.1
Vertebral fracture		469	4.8		3,154	7.7
Gait abnormality		660	6.7		4,126	10.1
HIV/AIDS		4	0.0		26	0.1
Hyperparathyroidism		76	0.8		434	1.1
Hyperthyroidism		2	0.0		4	0.0
Kyphosis		211	2.2		1,301	3.2
Liver disease		350	3.6		1,493	3.7
Osteoarthritis		4,129	42.0		18,247	44.6
Osteoporosis		5,448	55.4		24,651	60.2
Parkinson's disease		154	1.6		771	1.9
Chronic renal failure		369	3.8		1,959	4.8
Rheumatoid arthritis		583	5.9		2,796	6.8
Stroke or TIA		1,203	12.2		5,936	14.5
Syncope		714	7.3		3,505	8.6

Table continues

change in the PS is not necessarily equal to a 1-unit change in the DRS.

The objective of this study was to compare different calipers when matching on the DRS. Using 3 empirical examples (raloxifene vs. alendronate in 1-year fracture risk; cyclooxygenase 2 (COX-2) inhibitors vs. nonselective nonsteroidal antiinflammatory drugs (ns-NSAIDs) in 6-month gastrointestinal bleeding; and simvastatin + ezetimibe vs. simvastatin in

a 6-month composite measure of cardiovascular outcomes), we compared matching on the natural scale of the DRS with matching on fractions of the standard deviation of the logit of the DRS. We also compared these calipers using the DRS directly and using the prognostic propensity score (PPS) (8, 10), which is a transformation of the DRS to the PS scale calculated by including the DRS as the sole predictor variable in a PS model.

Table 1. Continued

Covariate ^a	Raloxifene (n = 9,829)			Alendronate (n = 40,960)		
	Mean (SD)	No.	%	Mean (SD)	No.	%
Prior medication use						
Alzheimer's drugs		303	3.0		1,503	3.7
Anticonvulsants		491	5.0		2,393	5.8
Non-SSRI antidepressants		1,122	11.4		4,256	10.4
SSRIs		1,416	14.4		5,819	14.2
Antipsychotic agents		284	2.9		1,343	3.3
β blockers		3,148	32.0		14,028	34.3
Benzodiazepines		2,643	26.9		10,107	24.7
Calcitonin		1,105	11.2		3,543	8.7
Bisphosphonates		36	0.4		143	0.4
Risedronate		380	3.9		1,192	2.9
Teriparatide		10	0.1		30	0.1
Corticosteroids		1,113	11.3		5,975	14.6
COX-2 inhibitors		2,190	22.3		9,841	24.0
Glitazones		363	3.7		1,817	4.4
Other diabetes drugs		1,172	11.9		5,488	13.4
Diuretics		1,040	10.6		4,945	12.1
Gastroprotective drugs		4,043	41.1		14,039	34.3
Hormone replacement therapy		1,724	17.5		3,638	8.9
Nonbenzodiazepine hypnotic agents		1,061	10.8		4,195	10.2
ns-NSAIDs		2,350	23.9		8,933	21.8
Parkinson's drugs		166	1.7		763	1.9
Thyroid hormone replacement		1,851	18.8		7,766	19.0

Abbreviations: AIDS, acquired immunodeficiency syndrome; COPD, chronic obstructive pulmonary disease; COX-2, cyclooxygenase 2; HIV, human immunodeficiency virus; ns-NSAID, nonselective nonsteroidal antiinflammatory drug; SD, standard deviation; SSRI, selective serotonin reuptake inhibitor; TIA, transient ischemic attack.

^a Covariates were assessed during a 180-day baseline period.

^b The combined comorbidity score is a combination of the Charlson and Elixhauser comorbidity scores (23).

METHODS

Databases

The study populations for the raloxifene and simvastatin + ezetimibe cohorts were drawn from a database of Medicare beneficiaries who were enrolled in state pharmaceutical benefits programs in Pennsylvania and New Jersey between 1994 and 2005. The study population for the COX-2 inhibitor cohort was drawn from Medicare patients enrolled in the Pennsylvania Pharmaceutical Assistance Contract (PACE) program between 1999 and 2002. These state programs provide medications at a reduced cost to low-income elderly persons who do not qualify for Medicaid. In each example, records of pharmacy dispensings were used to determine exposure, and linked Medicare information was used to assess covariates and outcomes.

Raloxifene and nonvertebral fracture cohort

We identified a cohort of new users of raloxifene or alendronate who initiated treatment between January 12, 1998, the first date on which new users of both medications appeared

in the data set, and December 31, 2005. New use was defined as not having a pharmacy dispensing for either raloxifene or alendronate in the 180 days prior to the date of the first eligible prescription (the index date). All cohort members must have had at least 180 days of continuous enrollment in the database prior to cohort entry, during which there must have been at least 1 prescription claim and 1 medical claim. Patients were excluded if they had a diagnosis of or treatment for Paget's disease (*International Classification of Diseases, Ninth Revision (ICD-9)*, code 731.0x) during the baseline period. Patients with dispensings for both raloxifene and alendronate on the index date were excluded. Exposure was classified according to the index drug for the duration of follow-up. The outcome of interest was first fracture of the hip, forearm, humerus, or pelvis within 365 days of the index date. The outcome was defined using ICD-9 diagnoses and Current Procedural Terminology, Fourth Edition, procedure codes and has been shown to have an overall positive predictive value of 94% in Medicare claims data (11). For a complete outcome definition, see the Web Appendix (available at <http://aje.oxfordjournals.org/>). Alendronate was chosen as the

comparator agent because it became available before raloxifene and was widely used in our data set.

We defined 57 covariates that were included in both PS and DRS models. The covariates were assessed in the 180 days prior to the index date and included information on demographic factors, health-care utilization, clinical conditions, and previous prescription drug use.

COX-2 inhibitors and gastrointestinal bleeding cohort

We identified new users of COX-2 inhibitors (rofecoxib or celecoxib) or ns-NSAIDs who initiated treatment between January 1, 1999, and December 31, 2002. New use was defined as no pharmacy dispensings for either exposure group during the 18 months prior to the index date. Patients were classified as exposed to their index drug for the duration of follow-up. The outcome was gastrointestinal bleeding within 180 days after treatment initiation, defined as a hospitalization for gastrointestinal hemorrhage or complications of peptic ulcer disease based on ICD-9 discharge diagnoses code 531.x, 532.x, 533.x, 534.x, 535.x, or 578.x. This definition has been shown to have a positive predictive value of 90% (12). We chose ns-NSAIDs as the reference group because they were available and widely used before COX-2 inhibitors entered the market (13).

The 18 covariates included in the PS and DRS models comprised demographic factors, health-care utilization variables, prior prescription drug use, and past clinical conditions.

Simvastatin + ezetimibe and cardiovascular outcomes cohort

We identified new users of a simvastatin + ezetimibe fixed-dose combination product and new users of simvastatin alone between August 12, 2004, the first date on which new users of both drugs were present in the data set, and December 31, 2005, the end of the study period. New use was defined as no pharmacy dispensing of either the index drug or the comparator drug during the 180 days prior to the index date. Patients were excluded if they used any lipid-lowering medication or received a diagnosis for any of the events included in the composite outcome during the baseline period. Exposure was classified according to index drug for the duration of follow-up. The outcome was a composite cardiovascular disease outcome comprising myocardial infarction, cerebrovascular events (subarachnoid or intracerebral hemorrhage, occlusion or stenosis of cerebral arteries, or acute cerebrovascular disease), acute coronary symptoms with revascularization, and death within 180 days of the index date. For a full definition of the outcome, see the Web Appendix (14, 15). Simvastatin was considered the reference group because it was available before simvastatin + ezetimibe combination products and was widely used in our data set.

The 63 covariates included in the PS and DRS models were assessed in the 180 days prior to treatment initiation and comprised demographic and health-care utilization variables, comorbid conditions, and prior drug use.

Statistical analysis

For each example, we computed 3 summary scores for each patient: a DRS, a PPS, and a PS. We used logistic regression

models to calculate PSs, which predicted the probability of receiving the exposure of interest (raloxifene, COX-2 inhibitors, simvastatin + ezetimibe) versus the referent (alendronate, ns-NSAIDs, and simvastatin, respectively), conditional on all baseline covariates. We also used logistic regression to calculate the DRSs, which predicted the probability of experiencing the outcome of interest during follow-up, conditional on the same variables. The DRS was estimated in the referent group, which has been shown to be less sensitive to modeling assumptions and to yield better balance than DRSs estimated in the full population in the presence of effect modifiers (7, 8, 16). The coefficients from the DRS models were then applied to estimate baseline disease risk for each member of the full study population. The PPS was estimated using a logistic model predicting the probability of initiation of the exposure of interest with a single linear term for the DRS as the sole predictor variable.

We used a nearest-neighbor algorithm to match patients on each score, without replacement, using 2 different matching ratios (17). In the first, up to 10 initiators of the referent drug with summary scores falling within the specified caliper were 1:N variable ratio-matched to each initiator of the drug of interest. Note that in the COX-2 inhibitor example, up to 10 initiators of COX-2 inhibitor use were matched to each ns-NSAID initiator, despite ns-NSAID users' being the reference group in the PS and odds ratio estimation processes, since COX-2 inhibitor use was much more common during the study period. Therefore, in this example, we estimated the average treatment effect among the untreated (i.e., among ns-NSAID initiators) instead of the average treatment effect in the treated, as in the other 2 examples. In the second matching strategy, initiators of the referent drug were 1:1 matched to initiators of the drug of interest. We performed matching using 8 different caliper widths: 0.3, 0.2, and 0.1 times the standard deviation of the logit of the summary score and 0.05, 0.025, 0.01, 0.001, and 0.0001 on the natural scale of the score.

After creating 1:N and 1:1 matched populations on all 3 summary scores using the calipers above, we calculated odds ratios and 95% confidence intervals comparing the drug of interest with the referent drug in each example. To each matched population, we fitted a conditional logistic regression model stratified by the matched set with exposure as the lone predictor variable. For each analysis, we also calculated the proportion of matched patients within each exposure group, as well as the proportion of the total fractures included in that matched analysis.

To assess caliper performance, we compared adjusted odds ratios after matching to each other as well as to the unadjusted association. Under the assumption that changes in the association were due to differences in confounding control, we considered calipers with adjusted odds ratios further from the unadjusted association to be less biased. While unmeasured confounding and differences in study populations and adherence preclude direct comparisons of associations with those from prior observational and randomized studies, these prior estimates can provide a rough guide about the direction and magnitude of the expected results. Thus, we expected the top-performing calipers to produce results close to the null in the raloxifene example, results slightly below the null in the

Table 2. Results From 1:N Variable Ratio Matching in the Example of Raloxifene Versus Alendronate Use and 1-Year Risk of Nonvertebral Fracture, United States, 1998–2005

Caliper Width	Odds of Fracture for Raloxifene Users vs. Alendronate Users ^a		Total No. of Patients Matched	Raloxifene Users Matched		Alendronate Users Matched		Fracture Events			
	OR	95% CI		No.	%	No.	%	Total No. of Fracture Events ^b	% of All Events Included ^c	No. of Events in Raloxifene Users	No. of Events in Alendronate Users
DRS											
0.3 × SD logit(DRS) ^d	1.01	0.90, 1.13	50,786	9,828	99.99	40,958	100.00	2,099	100.00	354	1,745
0.2 × SD logit(DRS)	1.00	0.90, 1.13	50,777	9,827	99.98	40,950	99.98	2,099	100.00	354	1,745
0.1 × SD logit(DRS)	1.00	0.89, 1.13	50,767	9,825	99.96	40,942	99.96	2,099	100.00	354	1,745
0.05	0.97	0.86, 1.08	50,788	9,829	100.00	40,959	100.00	2,099	100.00	354	1,745
0.025	1.00	0.89, 1.12	50,788	9,829	100.00	40,959	100.00	2,099	100.00	354	1,745
0.01	1.00	0.89, 1.13	50,770	9,829	100.00	40,941	99.95	2,098	99.95	354	1,744
0.001	1.00	0.89, 1.13	50,517	9,828	99.99	40,689	99.34	2,054	97.86	354	1,700
0.0001	1.02	0.90, 1.14	49,060	9,799	99.69	39,261	95.85	1,842	87.76	349	1,493
PPS											
0.3 × SD logit(PPS) ^d	1.00	0.90, 1.13	50,776	9,829	100.00	40,947	99.97	2,098	99.95	354	1,744
0.2 × SD logit(PPS)	1.00	0.89, 1.12	50,760	9,829	100.00	40,931	99.93	2,095	99.81	354	1,741
0.1 × SD logit(PPS)	1.00	0.89, 1.12	50,731	9,829	100.00	40,902	99.86	2,088	99.48	354	1,734
0.05	0.94	0.84, 1.06	50,789	9,829	100.00	40,960	100.00	2,099	100.00	354	1,745
0.025	0.99	0.89, 1.12	50,789	9,829	100.00	40,960	100.00	2,099	100.00	354	1,745
0.01	1.00	0.90, 1.13	50,788	9,829	100.00	40,959	100.00	2,099	100.00	354	1,745
0.001	1.01	0.90, 1.13	50,699	9,829	100.00	40,870	99.78	2,082	99.19	354	1,728
0.0001	1.02	0.91, 1.15	49,653	9,815	99.86	39,838	97.26	1,921	91.52	352	1,569
PS											
0.3 × SD logit(PS) ^d	1.02	0.91, 1.15	49,136	9,806	99.77	39,330	96.02	2,042	97.28	354	1,688
0.2 × SD logit(PS)	1.02	0.91, 1.15	49,125	9,802	99.73	39,323	96.00	2,042	97.28	354	1,688
0.1 × SD logit(PS)	1.02	0.91, 1.15	49,090	9,797	99.67	39,293	95.93	2,037	97.05	354	1,683
0.05	1.01	0.90, 1.14	49,412	9,806	99.77	39,606	96.69	2,052	97.76	354	1,698
0.025	1.02	0.91, 1.15	49,171	9,798	99.68	39,373	96.13	2,048	97.57	354	1,694
0.01	1.02	0.91, 1.15	49,124	9,794	99.64	39,330	96.02	2,042	97.28	354	1,688
0.001	1.02	0.91, 1.15	48,795	9,721	98.90	39,074	95.40	2,013	95.90	352	1,661
0.0001	1.02	0.91, 1.15	46,783	9,383	95.46	37,400	91.31	1,907	90.85	344	1,563

Abbreviations: CI, confidence interval; DRS, disease risk score; OR, odds ratio; PPS, prognostic propensity score; PS, propensity score; SD, standard deviation.

^a Unadjusted OR = 0.84 (95% CI: 0.75, 0.94).

^b All events included in the matched analysis of fracture risk. This was a subset of the total number of events in the cohort, as each matched analysis excluded persons who were unmatched, some of whom had the outcome of interest.

^c Percentage of the total number of events in the entire cohort (matched and unmatched) that were included in the matched analysis of fracture risk.

^d SD logit(DRS) = 0.8507; SD logit(PPS) = 0.2096; SD logit(PS) = 0.7049.

simvastatin + ezetimibe example, and results substantially below the null in the COX-2 inhibitor example (18–21).

RESULTS

Raloxifene and fracture example

We identified 9,829 new users of raloxifene and 40,960 new users of alendronate between 1998 and 2005. Over the

1-year follow-up period, there were 2,099 first fractures of the hip, forearm, humerus, or pelvis, for a cumulative 1-year incidence of 4.1%. Table 1 provides a comparison of baseline patient characteristics between raloxifene and alendronate initiators. Prior to matching, raloxifene users had fewer falls during the baseline period and less history of fractures of all types, and were less likely to have a baseline osteoporosis diagnosis. Histograms of the summary score distributions are displayed in Web Figures 1–3.

The unadjusted odds ratio for nonvertebral fracture was 0.84 (95% confidence interval (CI): 0.75, 0.94). When 1:*N* matching on the DRS, the 0.0001 caliper produced the odds ratio furthest from the crude estimate (odds ratio (OR) = 1.02, 95% CI: 0.90, 1.14), while the 0.05 caliper produced the odds ratio closest to the crude estimate (OR = 0.97, 95% CI: 0.86, 1.08) (Table 2). Similarly, when matching on the PPS, the 0.0001 caliper produced the odds ratio furthest from the crude estimate (OR = 1.02, 95% CI: 0.91, 1.15), while the 0.05 caliper was again closest to the unadjusted odds ratio (OR = 0.94, 95% CI: 0.84, 1.06). When 1:1 matching on the DRS or the PPS, the choice of caliper width did not change the association (Web Table 1).

COX-2 inhibitors and gastrointestinal bleeding example

We identified 32,042 COX-2 inhibitor initiators, 17,611 ns-NSAID initiators, and 552 occurrences of gastrointestinal bleeding within 180 days of treatment initiation between 1999 and 2002, for a cumulative incidence of 1.1%. Table 3 displays the baseline covariate balance between new users of COX-2 inhibitors and new users of ns-NSAIDs before matching. New users of COX-2 inhibitors had higher prevalences of previous gastrointestinal hemorrhage, use of gastroprotective drugs, and other comorbid conditions. Summary score distributions are presented in Web Figures 4–6.

The unadjusted odds ratio for gastrointestinal bleeding comparing COX-2 inhibitors with ns-NSAIDs was 1.09 (95% CI: 0.91, 1.30). After 1:*N* matching on the DRS, the odds ratio furthest from the crude estimate was 0.96 (95% CI: 0.80, 1.15) when matching within a 0.01 or 0.001 caliper, while the odds ratio closest to the crude estimate was 1.02 (95% CI: 0.85, 1.21) with a 0.05 caliper (Table 4). When using 1:*N*PPS matching, the odds ratio furthest from the unadjusted odds ratio was 0.95 (95% CI: 0.80, 1.14) when matching within the 0.025 caliper, while the odds ratio closest to the crude estimate was 1.01 (95% CI: 0.84, 1.21) with a 0.0001 caliper. As in the previous example, 1:1 matching was less sensitive to the choice of caliper width (Web Table 2).

Simvastatin + ezetimibe and cardiovascular outcomes cohort

We identified 1,976 new users of simvastatin + ezetimibe and 5,162 new users of simvastatin between 2004 and 2005. Within 180 days of treatment initiation, there were 1,252 occurrences of the composite outcome, for a cumulative incidence of 17.5%. Table 5 displays the covariate balance prior to matching. On average, new users of simvastatin + ezetimibe were younger, more likely to be female, and more likely to have had a cardiogram or diagnosis of hyperlipidemia during the baseline period prior to matching. Due to the higher outcome incidence in this example, the DRS distributions had larger mean values and variances. The histograms for each score distribution are presented in Web Figures 7–9.

The unadjusted odds ratio for the composite outcome comparing simvastatin + ezetimibe with simvastatin alone was 0.58 (95% CI: 0.50, 0.68). After 1:*N* matching on the DRS, the odds ratio furthest from the unadjusted estimate was 0.84 (95% CI: 0.70, 1.00) when matching within the 0.0001

Table 3. Baseline Characteristics of Initiators of COX-2 Inhibitor and ns-NSAID Use (Medicare Claims Data), United States, 1999–2002

Covariate ^a	COX-2 Inhibitors (n = 35,575)		ns-NSAIDs (n = 14,078)	
	No.	%	No.	%
Demographic factors				
Age ≥75 years	24,079	75.2	11,496	65.3
Female sex	27,528	85.9	14,293	81.2
White race/ethnicity	30,583	95.5	15,808	89.8
Utilization of health services				
>4 distinct generic drugs in previous year	24,120	75.3	11,852	67.3
>4 physician visits in previous year	22,919	71.5	11,363	64.5
Hospitalized in previous year	9,804	30.6	4,591	26.1
Nursing home resident	2,671	8.3	996	5.7
Prior medication use				
Gastroprotective drugs	8,785	27.4	3,600	20.4
Warfarin	4,252	13.3	1,153	6.6
Corticosteroids	2,800	8.7	1,373	7.8
Clinical conditions				
Charlson comorbidity score ≥1	24,343	76.0	12,521	71.1
History of osteoarthritis	15,549	48.5	5,898	33.5
History of rheumatoid arthritis	1,602	5.0	476	2.7
History of peptic ulcers	1,189	3.7	426	2.4
History of gastrointestinal hemorrhage	551	1.7	196	1.1
History of hypertension	23,332	72.8	12,363	70.2
History of congestive heart failure	9,727	30.4	4,328	24.6
History of coronary artery disease	5,266	16.4	2,603	14.8

Abbreviations: COX-2, cyclooxygenase 2; ns-NSAID, nonselective nonsteroidal antiinflammatory drug.

^a Covariates were assessed during a 180-day baseline period.

caliper (Table 6). The odds ratio closest to the unadjusted estimate was 0.78 (95% CI: 0.68, 0.90) when using any logit-based caliper or a natural-scale caliper of 0.025 or larger. When matching 1:*N* on the PPS, the odds ratio closest to the unadjusted estimate was 0.78 (95% CI: 0.68, 0.90) using several calipers, while the odds ratio furthest from the unadjusted estimate was 0.80 (95% CI: 0.68, 0.93) with a caliper of 0.0001. The 1:1 matching results were almost identical to the 1:*N* matching results and are displayed in Web Table 3.

DISCUSSION

When 1:*N* variable-ratio matching on a DRS using an optimal nearest-neighbor matching algorithm, we found that natural-scale calipers commonly used for PS matching (e.g., 0.05) may be too large for DRS matching when the outcome is uncommon. In the raloxifene and COX-2 inhibitor

Table 4. Results From 1:N Variable Ratio Matching in the Example of COX-2 Versus ns-NSAID Use and Gastrointestinal Bleeding, United States, 1999–2002

Caliper Width	Odds of GI Bleeding for COX-2 Inhibitor Users vs. ns-NSAID Users ^a		Total No. of Patients Matched	COX-2 Inhibitor Users Matched		ns-NSAID Users Matched		GI Bleeding Events			
	OR	95% CI		No.	%	No.	%	Total No. of GI Bleeding Events ^b	% of All Events Included ^c	No. of Events in COX-2 Inhibitor Users	No. of Events in ns-NSAID Users
DRS											
0.3 × SD logit(DRS) ^d	0.97	0.81, 1.16	46,888	30,073	93.85	16,815	95.48	532	96.38	354	178
0.2 × SD logit(DRS)	0.97	0.81, 1.16	46,887	30,072	93.85	16,815	95.48	532	96.38	354	178
0.1 × SD logit(DRS)	0.97	0.81, 1.16	46,882	30,069	93.84	16,813	95.47	532	96.38	354	178
0.05	1.02	0.85, 1.21	49,640	32,030	99.96	17,610	99.99	552	100.00	367	185
0.025	0.99	0.83, 1.18	49,640	32,030	99.96	17,610	99.99	552	100.00	367	185
0.01	0.96	0.80, 1.15	49,639	32,029	99.96	17,610	99.99	552	100.00	367	185
0.001	0.96	0.80, 1.15	49,599	31,996	99.86	17,603	99.95	551	99.82	366	185
0.0001	0.97	0.81, 1.17	49,344	31,848	99.39	17,496	99.35	544	98.55	363	181
PPS											
0.3 × SD logit(PPS) ^d	0.96	0.80, 1.14	49,286	31,757	99.11	17,529	99.53	546	98.91	363	183
0.2 × SD logit(PPS)	0.96	0.80, 1.15	49,280	31,752	99.09	17,528	99.53	545	98.73	362	183
0.1 × SD logit(PPS)	0.96	0.80, 1.15	49,246	31,728	99.02	17,518	99.47	545	98.73	362	183
0.05	0.96	0.80, 1.15	49,238	31,726	99.01	17,512	99.44	546	98.91	363	183
0.025	0.95	0.80, 1.14	49,238	31,726	99.01	17,512	99.44	546	98.91	363	183
0.01	0.96	0.80, 1.15	49,237	31,725	99.01	17,512	99.44	546	98.91	363	183
0.001	0.97	0.81, 1.16	49,131	31,661	98.81	17,470	99.20	542	98.19	361	181
0.0001	1.01	0.84, 1.21	48,553	31,445	98.14	17,088	97.03	521	94.38	348	173
PS											
0.3 × SD logit(PS) ^d	0.92	0.76, 1.10	48,143	31,700	98.93	16,443	93.37	542	98.19	363	179
0.2 × SD logit(PS)	0.91	0.76, 1.10	48,126	31,699	98.93	16,427	93.28	542	98.19	363	179
0.1 × SD logit(PS)	0.91	0.76, 1.09	48,086	31,693	98.91	16,393	93.08	540	97.83	362	178
0.05	0.93	0.77, 1.12	48,156	31,735	99.04	16,421	93.24	540	97.83	363	177
0.025	0.92	0.77, 1.11	48,111	31,733	99.04	16,378	93.00	539	97.64	363	176
0.01	0.92	0.76, 1.10	48,099	31,727	99.02	16,372	92.96	539	97.64	363	176
0.001	0.92	0.76, 1.11	47,937	31,681	98.87	16,256	92.31	537	97.28	361	176
0.0001	0.96	0.79, 1.16	45,777	30,294	94.54	15,483	87.92	502	90.94	338	164

Abbreviations: CI, confidence interval; COX-2, cyclooxygenase 2; DRS, disease risk score; GI, gastrointestinal; ns-NSAID, nonselective nonsteroidal antiinflammatory drug; OR, odds ratio; PPS, prognostic propensity score; PS, propensity score; SD, standard deviation.

^a Unadjusted OR = 1.09 (95% CI: 0.91, 1.30).

^b All events included in the matched analysis of GI bleeding. This was a subset of the total number of events in the cohort, as each matched analysis excluded persons who were unmatched, some of whom had the outcome of interest.

^c Percentage of the total number of events in the entire cohort (matched and unmatched) that were included in the matched analysis of GI bleeding.

^d SD logit(DRS) = 0.7053; SD logit(PPS) = 0.2090; SD logit(PS) = 0.5896.

examples, with cumulative outcome incidences of 4% and 1%, respectively, a caliper of 0.05 on the natural scale encompassed nearly all of the DRS distributions. However, in the simvastatin + ezetimibe example, with a cumulative outcome incidence of 17%, the 0.05 caliper performed similarly to the other calipers. In general, matching on the DRS may require finer calipers than matching on the PS, because a difference in baseline outcome probabilities between exposure groups is

the very definition of confounding and a difference in estimated disease risks between treatment groups guarantees confounding provided that the DRS model is well specified. Differences in exposure probability between exposure groups would lead to confounding commensurate with the degree to which exposure probability is related to outcome risk.

Calipers smaller than 0.05 on the natural scale and calipers based on fractions of the standard deviation of the logit of

Table 5. Baseline Characteristics of Initiators of Simvastatin + Ezetimibe Use and Simvastatin Use (Medicare Claims Data), United States, 2004–2005

Covariate ^a	Simvastatin + Ezetimibe (n = 1,976)			Simvastatin (n = 5,162)		
	Mean (SD)	No.	%	Mean (SD)	No.	%
Demographic factors						
Age, years	75.6 (6.6)			76.2 (7.1)		
Female sex		1,550	78.4		3,767	73.0
White race/ethnicity		1,760	89.1		4,616	89.4
Utilization of health services						
No. of physician visits	4.9 (3.7)			4.7 (4.0)		
No. of cardiovascular physician visits	2.5 (2.3)			2.2 (2.3)		
No. of cardiovascular diagnoses	4.0 (4.0)			4.3 (5.0)		
No. of hospital admissions	0.1 (0.4)			0.2 (0.6)		
No. of days hospitalized	0.5 (2.6)			1.8 (6.3)		
No. of cardiovascular hospital admissions	0.04 (0.2)			0.1 (0.4)		
No. of days in cardiovascular hospital	0.2 (1.4)			0.8 (3.6)		
No. of days in nursing home	0.3 (3.2)			1.1 (6.8)		
No. of distinct generic drugs	7.5 (4.0)			7.7 (4.3)		
Bone mineral density testing		103	5.2		254	4.9
Lipid testing		722	36.5		1,984	38.4
Cardiogram		1,044	52.8		1,941	37.6
Preventive care ^b		631	31.9		1,491	28.9
Clinical conditions						
Combined comorbidity score ^c	0.6 (1.8)			0.9 (2.1)		
Peripheral vascular disease		182	9.1		583	11.3
Diabetes mellitus		813	41.1		1,954	37.9
Hyperlipidemia		1,662	84.1		3,740	72.5
CABG prior to baseline period		46	2.3		196	3.8
CABG during baseline period		4	0.2		17	0.3
Hypertension		1,550	78.4		3,803	73.7
Congestive heart failure (any diagnosis)		205	10.4		706	13.7
Congestive heart failure (hospital diagnosis)		16	0.8		151	2.9
Atrial fibrillation (hospital diagnosis)		17	0.9		131	2.5
Chronic obstructive pulmonary disease		276	14.0		813	15.8
Chest pain		284	14.4		837	16.2
Coronary atherosclerosis		492	24.9		1,429	27.7
Conduct disorder		47	2.4		188	3.6

Table continues

the DRS performed relatively well in all examples. While natural-scale calipers are most commonly reported in the PS literature, Cochran and Rubin (22) provided the theoretical rationale that matching on a continuous, normally distributed variable using a caliper of 0.2 times the standard deviation of that variable removes more than 99% of bias. In a simulation study, Austin (5) showed that PS calipers between 0.005 and 0.03 on the natural scale reduced bias more than a caliper of 0.2 times the standard deviation of the logit of the PS but that the latter yielded lower mean squared error by including more matched pairs in the analysis. A practical advantage of

logit-based calipers appears to be that they are less sensitive than natural-scale calipers to the score distributions.

Matching on the PPS appeared to offer no consistent practical advantage over matching on the DRS directly. In the raloxifene example, PPS matching using the 0.05 caliper produced estimates closer to the crude estimate than matching on the DRS. Hansen (8) originally proposed the PPS as a way to reduce bias in estimated treatment effects by balancing treatment groups on prognostically relevant variables, and Leacy and Stuart (10) suggested the use of ordinary PS calipers when matching on the PPS. More empirical work is needed

Table 5. Continued

Covariate ^a	Simvastatin + Ezetimibe (n = 1,976)			Simvastatin (n = 5,162)		
	Mean (SD)	No.	%	Mean (SD)	No.	%
Heart palpitations		87	4.4		177	3.4
Ischemic heart disease		537	27.2		1,547	30.0
Alzheimer's disease		62	3.1		242	4.7
Cancer		268	13.6		772	15.0
Depression		121	6.1		401	7.8
Falls		12	0.6		62	1.2
Hip fracture		10	0.5		48	0.9
Hyperparathyroidism		6	0.3		23	0.5
Osteoarthritis		463	23.4		1,083	20.9
Osteoporosis		225	11.4		564	10.9
Chronic renal disease		71	3.6		265	5.1
End-stage renal disease		2	0.1		4	0.1
Rheumatoid arthritis		53	2.7		110	2.1
Urinary tract infection		182	9.2		510	9.9
Prior medication use						
ACE inhibitors		430	21.8		1,230	23.8
α blockers		17	0.9		72	1.4
Antiarrhythmic agents		50	2.5		148	2.9
Antifungal agents		33	1.7		78	1.5
Angiotensin receptor blockers		252	12.8		587	11.4
β blockers		697	35.3		1,828	35.4
Calcium channel blockers		479	24.2		1,309	25.4
Diabetes drugs		530	26.8		1,336	25.9
Erectile drugs		15	0.8		45	0.9
Hormone replacement therapy		69	3.5		133	2.6
Loop diuretics		325	16.5		911	17.7
Nonsteroidal antiinflammatory drugs		383	19.4		886	17.2
Osteoporosis drugs		281	14.2		760	14.7
Potassium-sparing agents/aldosterone		74	3.7		204	4.0
Proton pump inhibitors		454	23.0		1,087	21.1
Psychoactive agents		613	31.0		1,572	30.5
Thiazides		228	11.5		654	12.7
Warfarin		159	8.1		488	9.5

Abbreviations: ACE, angiotensin-converting enzyme; CABG, coronary artery bypass grafting; SD, standard deviation.

^a Covariates were assessed during a 180-day baseline period.

^b Gynecological examination, prophylactic vaccination, routine medical examination, or screening mammogram.

^c The combined comorbidity score is a combination of the Charlson and Elixhauser comorbidity scores (23).

to determine whether and when matching on the PPS has any advantage over matching on the DRS.

When we matched 1:1 on any summary score, the choice of caliper width had a relatively small impact on the associations in all examples. Because of the strong overlap in summary score distributions in our studies and the use of the nearest-neighbor matching algorithm, the single closest match for a given patient was likely to have been well within any reasonable

caliper. This is in contrast to the results of the primary variable ratio matched analyses, which depended more heavily on the specified caliper. For example, in the raloxifene analysis, the average difference in DRSs among the top 10% of matched pairs with the largest differences was 2.44×10^{-5} when matching at a 1:1 ratio using a natural-scale caliper of 0.05 as compared with 0.042 for variable ratio matching using the same caliper. Because variable ratio matching finds

Table 6. Results From 1:N Variable Ratio Matching in the Example of Simvastatin + Ezetimibe Use Versus Simvastatin Use and 6-Month Cardiovascular Outcomes,^a United States, 2004–2005

Caliper Width	Odds of a CVD Event for Simvastatin + Ezetimibe Users vs. Simvastatin Users ^b		Total No. of Patients Matched	Simvastatin + Ezetimibe Users Matched		Simvastatin Users Matched		CVD Events			
	OR	95% CI		No.	%	No.	%	Total No. of CVD Events ^c	% of All Events Included ^d	No. of Events in Simvastatin + Ezetimibe Users	No. of Events in Simvastatin Users
DRS											
0.3 × SD logit(DRS) ^e	0.78	0.68, 0.90	7,129	1,976	99.83	5,153	99.36	1,244	99.36	245	999
0.2 × SD logit(DRS)	0.78	0.68, 0.90	7,127	1,976	99.79	5,151	99.28	1,243	99.28	245	998
0.1 × SD logit(DRS)	0.78	0.68, 0.90	7,116	1,976	99.57	5,140	98.72	1,236	98.72	245	991
0.05	0.78	0.68, 0.90	7,135	1,976	99.94	5,159	99.76	1,249	99.76	245	1,004
0.025	0.78	0.68, 0.90	7,128	1,976	99.81	5,152	99.28	1,243	99.28	245	998
0.01	0.79	0.68, 0.91	7,110	1,976	99.46	5,134	98.08	1,228	98.08	245	983
0.001	0.80	0.69, 0.93	6,730	1,954	92.52	4,776	81.79	1,024	81.79	236	788
0.0001	0.84	0.70, 1.00	4,761	1,603	61.18	3,158	45.93	575	45.93	178	397
PPS											
0.3 × SD logit(PPS) ^e	0.78	0.68, 0.90	7,135	1,976	99.94	5,159	99.76	1,249	99.76	245	1,004
0.2 × SD logit(PPS)	0.78	0.68, 0.90	7,133	1,976	99.90	5,157	99.60	1,247	99.60	245	1,002
0.1 × SD logit(PPS)	0.79	0.68, 0.90	7,123	1,976	99.71	5,147	99.04	1,240	99.04	245	995
0.05	0.78	0.68, 0.90	7,138	1,976	100.00	5,162	100.00	1,252	100.00	245	1,007
0.025	0.78	0.68, 0.90	7,138	1,976	100.00	5,162	100.00	1,252	100.00	245	1,007
0.01	0.78	0.68, 0.90	7,136	1,976	99.96	5,160	99.84	1,250	99.84	245	1,005
0.001	0.79	0.68, 0.91	7,054	1,974	98.41	5,080	95.69	1,198	95.69	245	953
0.0001	0.80	0.68, 0.93	6,129	1,870	82.51	4,259	67.09	840	67.09	219	621
PS											
0.3 × SD logit(PS) ^e	0.78	0.68, 0.90	7,017	1,967	97.83	5,050	95.37	1,194	95.37	244	950
0.2 × SD logit(PS)	0.78	0.68, 0.90	6,999	1,961	97.60	5,038	94.89	1,188	94.89	243	945
0.1 × SD logit(PS)	0.78	0.68, 0.90	6,964	1,956	97.02	5,008	94.01	1,177	94.01	243	934
0.05	0.77	0.67, 0.89	7,057	1,967	98.61	5,090	96.81	1,212	96.81	244	968
0.025	0.78	0.67, 0.90	7,018	1,959	98.00	5,059	95.69	1,198	95.69	243	955
0.01	0.78	0.68, 0.90	6,994	1,954	97.64	5,040	95.21	1,192	95.21	243	949
0.001	0.78	0.68, 0.91	6,763	1,901	94.19	4,862	89.30	1,118	89.30	240	878
0.0001	0.84	0.71, 1.00	4,504	1,494	58.31	3,010	54.47	682	54.47	197	485

Abbreviations: CI, confidence interval; CVD, cardiovascular disease; DRS, disease risk score; OR, odds ratio; PPS, prognostic propensity score; PS, propensity score; SD, standard deviation.

^a The outcome was a composite CVD measure comprising myocardial infarction, cerebrovascular events (subarachnoid or intracerebral hemorrhage, occlusion or stenosis of cerebral arteries, or acute cerebrovascular disease), acute coronary symptoms with revascularization, and death within 180 days of the index date.

^b Unadjusted OR = 0.58 (95% CI: 0.50, 0.68).

^c All events included in the matched analysis of CVD events. This was a subset of the total number of events in the cohort, as each matched analysis excluded persons who were unmatched, some of whom had the outcome of interest.

^d Percentage of the total number of events in the entire cohort (matched and unmatched) that were included in the matched analysis of CVD events.

^e SD logit(DRS) = 0.9901; SD logit(PPS) = 0.2724; SD logit(PS) = 0.6917.

all acceptable matches within the caliper, more matches occur at the edge of the caliper.

Our results should be interpreted in the context of limitations of these analyses. We lacked a true gold standard with which to compare our associations. Therefore, we used differences among adjusted and unadjusted associations to assess the relative performance of the different calipers. We assumed that changes in estimates were due to differences in confounding. However, different matching strategies yield different matched populations, which could result in different associations independent of confounding. We sought results of previous randomized trials as rough guides about the direction and magnitude of the expected results, but these trials do not necessarily provide accurate estimates of the true treatment effects in our observational cohorts. Simulation studies are needed to precisely quantify the amount of bias associated with using certain calipers to match on the DRS and the PPS in different scenarios with varying degrees of confounding, summary score distribution overlap, relative exposure group size, and outcome incidences. However, the generalizability of our findings to situations with less overlap in the DRS distributions is evidenced by the fact that the smallest caliper of 0.0001 matched only 61% of exposed patients in the simvastatin + ezetimibe example, while in each of the other examples at least 95% of exposed patients were matched using the same caliper. Finally, previous comparisons of PS calipers have used covariate balance as a metric to assess relative performance. However, because DRS matching balances baseline disease risk and not necessarily covariate distributions, we were not able to use covariate balance metrics.

In conclusion, when we employed 1:N matching on the DRS in settings with uncommon outcomes, certain commonly used PS calipers on the natural scale were too wide and produced estimates which were probably biased. When outcomes are common, all commonly used calipers appear to perform similarly well for DRS matching. Using calipers based on a logit transformation or using natural-scale calipers smaller than 0.05 may be advisable for DRS matching in general, but simulation studies are needed to identify the optimal caliper width. We also found that the PPS may serve as a valid method for matching indirectly on the DRS in certain situations, but more work is necessary to elucidate its practical advantages.

ACKNOWLEDGMENTS

Author affiliations: Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts (John G. Connolly, Joshua J. Gagne).

Both authors contributed equally to the work.

This work was supported by a KL2/Catalyst Medical Research Investigator Training (CMerIT) award (an appointed KL2 award) from Harvard Catalyst | The Harvard Clinical and Translational Science Center (National Center for Research Resources and National Center for Advancing Translational Sciences award KL2 TR001100).

The content of this article is solely the responsibility of the authors and does not necessarily represent the official views of

Harvard Catalyst, Harvard University and its affiliated academic health-care centers, or the National Institutes of Health.

Conflict of interest: none declared.

REFERENCES

- Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983;70(1):41–55.
- Austin PC. A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003. *Stat Med*. 2008;27(12):2037–2049.
- Austin PC. Primer on statistical interpretation or methods report card on propensity-score matching in the cardiology literature from 2004 to 2006: a systematic review. *Circ Cardiovasc Qual Outcomes*. 2008;1(1):62–67.
- Rosenbaum PR, Rubin DB. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *Am Stat*. 1985;39(1):33–38.
- Austin PC. Some methods of propensity-score matching had superior performance to others: results of an empirical investigation and Monte Carlo simulations. *Biom J*. 2009;51(1):171–184.
- Cadarette SM, Gagne JJ, Solomon DH, et al. Confounder summary scores when comparing the effects of multiple drug exposures. *Pharmacoepidemiol Drug Saf*. 2010;19(1):2–9.
- Glynn RJ, Gagne JJ, Schneeweiss S. Role of disease risk scores in comparative effectiveness research with emerging therapies. *Pharmacoepidemiol Drug Saf*. 2012;21(suppl 2):138–147.
- Hansen BB. The prognostic analogue of the propensity score. *Biometrika*. 2008;95(2):481–488.
- Wyss R, Ellis AR, Brookhart MA, et al. Matching on the disease risk score in comparative effectiveness research of new treatments. *Pharmacoepidemiol Drug Saf*. 2015;24(9):951–961.
- Leacy FP, Stuart EA. On the joint use of propensity and prognostic scores in estimation of the average treatment effect on the treated: a simulation study. *Stat Med*. 2014;33(20):3488–3508.
- Ray WA, Griffin MR, Fought RL, et al. Identification of fractures from computerized Medicare files. *J Clin Epidemiol*. 1992;45(7):703–714.
- Raiford DS, Pérez Gutthann S, García Rodríguez LA. Positive predictive value of ICD-9 codes in the identification of cases of complicated peptic ulcer disease in the Saskatchewan hospital automated database. *Epidemiology*. 1996;7(1):101–104.
- Moore RA, Derry S, Makinson GT, et al. Tolerability and adverse events in clinical trials of celecoxib in osteoarthritis and rheumatoid arthritis: systematic review and meta-analysis of information from company clinical trial reports. *Arthritis Res Ther*. 2005;7(3):R644–R665.
- Kiyota Y, Schneeweiss S, Glynn RJ, et al. Accuracy of Medicare claims-based diagnosis of acute myocardial infarction: estimating positive predictive value on the basis of review of hospital records. *Am Heart J*. 2004;148(1):99–104.
- Tirschwell DL, Longstreth WT Jr. Validating administrative data in stroke research. *Stroke*. 2002;33(10):2465–2470.
- Arbogast PG, Ray WA. Performance of disease risk scores, propensity scores, and traditional multivariable outcome regression in the presence of multiple confounders. *Am J Epidemiol*. 2011;174(5):613–620.
- Rassen JA, Doherty M, Huang W, et al. Pharmacoepidemiology toolbox including high-dimensional propensity score (hd-PS) adjustment version 2. http://www.drugapi.org/dope-downloads/#Pharmacoepidemiology_Toolbox. 2011. Accessed June 6, 2015.

18. Lin T, Yan SG, Cai XZ, et al. Alendronate versus raloxifene for postmenopausal women: a meta-analysis of seven head-to-head randomized controlled trials. *Int J Endocrinol*. 2014;2014:796510.
19. Schneeweiss S, Rassen JA, Glynn RJ, et al. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. *Epidemiology*. 2009;20(4):512–522.
20. Cannon CP, Blazing MA, Giugliano RP, et al. Ezetimibe added to statin therapy after acute coronary syndromes. *N Engl J Med*. 2015;372(25):2387–2397.
21. Bombardier C, Laine L, Reicin A, et al. Comparison of upper gastrointestinal toxicity of rofecoxib and naproxen in patients with rheumatoid arthritis. VIGOR Study Group. *N Engl J Med*. 2000;343(21):1520–1528.
22. Cochran WG, Rubin DB. Controlling bias in observational studies: a review. *Sankhyā*. 1973;35(4):417–446.
23. Gagne JJ, Glynn RJ, Avorn J, et al. A combined comorbidity score predicted mortality in elderly patients better than existing scores. *J Clin Epidemiol*. 2011;64(7):749–759.