PHARMACOEPIDEMIOLOGY AND DRUG SAFETY 2012; **21**(S2): 138–147
Published online in Wiley Online Library (wileyonlinelibrary.com) **DOI**: 10.1002/pds.3231

ORIGINAL REPORT

Role of disease risk scores in comparative effectiveness research with emerging therapies

Robert J. Glynn*, Joshua J. Gagne and Sebastian Schneeweiss

Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

ABSTRACT

Background Usefulness of propensity scores and regression models to balance potential confounders at treatment initiation may be limited for newly introduced therapies with evolving use patterns.

Objectives To consider settings in which the disease risk score has theoretical advantages as a balancing score in comparative effectiveness research because of stability of disease risk and the availability of ample historical data on outcomes in people treated before introduction of the new therapy.

Methods We review the indications for and balancing properties of disease risk scores in the setting of evolving therapies and discuss alternative approaches for estimation. We illustrate development of a disease risk score in the context of the introduction of atorvastatin and the use of high-dose statin therapy beginning in 1997, based on data from 5668 older survivors of myocardial infarction who filled a statin prescription within 30 days after discharge from 1995 until 2004. Theoretical considerations suggested development of a disease risk score among nonusers of atorvastatin and high-dose statins during the period 1995–1997.

Results Observed risk of events increased from 11% to 35% across quintiles of the disease risk score, which had a *C*-statistic of 0.71. The score allowed control of many potential confounders even during early follow-up with few study endpoints.

Conclusions Balancing on a disease risk score offers an attractive alternative to a propensity score in some settings such as newly marketed drugs and provides an important axis for evaluation of potential effect modification. Joint consideration of propensity and disease risk scores may be valuable. Copyright © 2012 John Wiley & Sons, Ltd.

KEY WORDS—confounding factors (epidemiology); propensity scores; epidemiological methods; statistical models

Received 16 August 2011; Revised 27 December 2011; Accepted 17 January 2012

INTRODUCTION

Special challenges apply to the control of confounding in studies of the safety and effectiveness of new and evolving therapies. Many covariates can influence choices among alternative therapies, and prescriber preferences often evolve quickly during the period of early experience with a specific drug or dose.[1] Especially in early follow-up, there are typically relatively few study outcomes.

This setting of evolving prescriber preferences and relatively few outcomes can limit the use of both traditional multivariable models and propensity scores as approaches to obtain unbiased estimates of relative treatment effects. Whether one uses a case–control or prospective study design to compare outcomes across treatment groups, the number of potential confounders included in a standard regression approach is limited by the number of study endpoints. For example, reliable estimation in both logistic regression and proportional hazards models requires no more than one covariate (counting separately interaction and higher-order terms) for every 10 study outcomes.[2] This can lead investigators to prioritize potential confounders and exclude some of theoretical relevance from multivariable analyses, leading to suboptimal confounder control.

Propensity scores are a valuable strategy to reduce the dimension of potential confounding variables and can be particularly useful when there are relatively few study endpoints.[3,4] However, with new and evolving therapies, a prescriber's preference regarding the

*Correspondence to: R. J. Glynn, Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, 1620 Tremont Street, Boston, MA 02120, USA. E-mail: rglynn@rics.bwh.harvard.edu

characteristics of patients that indicate a specific drug choice is likely to change, and at varying rates across providers. Patients may also have varying attitudes about use of new therapies. Such evolving relationships of specific characteristics with treatment choices imply absence of a sharply defined propensity score and can lead investigators to consideration of time-varying propensity scores.[5] In addition, if some variables are related to treatment choice but not study outcomes, these instruments are best not included in the propensity score,[6–8] but their identification in the setting of newly evolving therapies is challenging.

With these challenges to multivariable analysis and propensity score estimation, a disease risk score can be a useful tool for confounder control. Here, we provide some background on the use of the disease risk score in epidemiology, consider controversies regarding its estimation, note the balancing properties of this score, and illustrate the development of a disease risk score with examples from the use of statins, including high-intensity statin regimens, after myocardial infarction.

OVERVIEW OF DISEASE RISK SCORES

Whereas a propensity score summarizes the way potential risk factors differ between users of alternative treatments to be compared, a disease risk score characterizes the relationship of risk factors with the study outcome. Summary measures of disease risk play a prominent role in guidelines for drug use and the evaluation of possible effect measure modification of new treatments or new indications for established therapies. For example, 5-year risk of invasive breast cancer estimated from the Gail model critically influences use of selective estrogen receptor modulators for risk reduction.[9,10] Similarly, guidelines for the use of statins in the primary prevention of cardiovascular disease incorporate the Framingham risk score as a key determinant of treatment eligibility.[11] Other summary measures of disease severity that often direct treatments include the Acute Physiology, Age, Chronic Health Evaluation (APACHE) III score in intensive care patients,[12] the National Institutes of Health stroke scale,[13] and the Glasgow coma scale.[14] Strengths of these scales include their applicability in different populations and periods.

When summary evidence suggests the value of a treatment in a target population, a disease risk score provides an important axis for evaluation of possibly varying effects and for characterization of subgroup-specific absolute treatment effects. For example, in consideration of the use of statins for primary prevention, treatment guidelines require specific information on risks and benefits within categories of absolute disease risk.[15]

Although disease risk scores with prespecified weights are a useful tool for confounder adjustment in studies of treatment effectiveness and safety, they are seldom sufficient to completely control for potential confounding. In the use of administrative data, the predictive ability of available comorbidity scores such as those developed by Charlson *et al.*[16] and Elixhauser *et al.*[17] can be enhanced through estimation of weights for their components within the study population of interest.[18,19] However, these risk scores are generally considered to be only one component of a strategy to control confounding, rather than a self-sufficient approach. Even when total mortality is the study endpoint, administrative datasets generally contain additional determinants of death that are not included in available scores. A wider view of such potential determinants is generally required for adequate confounder adjustment, compared with the perspective provided by construction of a parsimonious comorbidity index. As in the construction of a propensity score,[20] the disease risk score should err on the side of inclusion of variables that show even a modest association with the outcome.

The notion that a study-specific disease risk score alone can control confounding and aid in causal inference has a substantial history. Peters[21] and Belson[22] proposed a two-step approach for confounder adjustment with the first-stage development of a model to predict the outcome among the unexposed, followed by adjustment for the predicted outcome in a comparison between the exposed and unexposed. Cochran[23] described the conditions under which this Peters–Belson approach is preferable to multivariable adjustment for causal inference. In particular, this approach has theoretical advantages in the presence of effect measure modification across the dimension of outcome risk in the unexposed and has extensive applications in economics and health services research.[24]

In considering the value of alternative summary confounder scores to reveal potential problems with a multivariable analysis of the effects of an exposure, Miettinen recommended the use of a form of disease risk score.[25] Specifically, in the setting of a case–control study, he recommended inclusion of the exposure status and all potential confounding variables in a multivariable model to predict the study outcome. Then, each subject's predicted risk was obtained by setting the exposure status to zero, and stratified analysis was used to evaluate the relationship of the exposure and outcome.

DOI: 10.1002/pds

An important theoretical development in understanding the disease risk score is an appreciation of its balancing property as described by Hansen, which parallels the balancing property of the propensity score.[26] Specifically, with a properly developed propensity score $PS(X)$ to summarize the way a vector of covariates X predicts treatment assignment, Rosenbaum and Rubin showed that if sufficiently large groups of exposed and unexposed subjects with the same value of $PS(X)$ are identified, these two groups will have the same distributions of all components of X.[27,28] This implies that stratification or matching on the propensity score can yield a better exposed/unexposed balance of these measured covariates than would be obtained by randomized treatment assignment.[29]

In parallel, a well-formed disease risk score $DR(X)$ has the property that the potential outcome if untreated is independent of covariates X, given $DR(X)$. Note that this is a balance of disease risks, as distinct from the balance of treatment propensities provided by the propensity score. This prognostic balance can only be evaluated in the untreated. Furthermore, as Hansen pointed out, its evaluation in the untreated subjects within a population including treated and untreated subjects requires an assumption: that the potential outcome if untreated is independent of the actual treatment assignment given X.[26] This is an assumption of no unmeasured confounders outside X. Table 1 summarizes the aspects of study design that can influence the relative utility of disease risk scores and propensity scores.

ALTERNATIVE ESTIMATION STRATEGIES FOR THE DISEASE RISK SCORE

The above discussion indicates use of three distinct populations to develop a disease risk score: (i) in an alternative data set or in a period prior to the current study, perhaps before introduction of a new therapy; (ii) in the study population, but based on estimation of disease risk in the unexposed group only, akin to the Peters–Belson method; and (iii) in the entire study population, based on a model including indicators of exposure status, and then set this exposure status to zero for an individual's predicted risk, as suggested by Miettinen. Each approach seeks to estimate a disease risk score that will be the most representative of the study population, and each has both strengths and limitations.

Estimation of a disease risk score using all subjects in the study population, based on a model with an indicator for exposure status, benefits from the ready availability of the data set and its use of a larger sample size than does estimation restricted to the unexposed, to yield potentially more reliable estimates of disease risk under the assumption of a correct model form. Several simulation studies have found that stratification on a disease risk score obtained in this way (according to the suggestion of Miettinen) performs comparably to both propensity score stratification and multivariable analysis, as long as covariates are not too highly correlated with exposure.[30–33] Further within the context of the scenarios examined, this full-cohort disease risk score can sometimes outperform a

Table 1. Study design features that influence the value or feasibility of disease risk scores (DRS) or propensity scores (PS)

Study feature or analytic goal	Impact on DRS	Impact on PS
Ample historical data (before new treatment)	Very useful for DRS development	Informs variable selection, but not generally used in estimation
Rare outcome	Greatly limits DRS development and usefulness	PS particularly valuable, but limited ability to exclude possible instrumental variables
Rare exposure	Little impact on DRS	Limits estimation of PS
Rapidly evolving treatment indications	Little impact on DRS	Challenges ability to fit PS and suggests time interactions or time-specific PS
Interest in >1 outcome/>2 exposures	DRS may be particularly useful with >2 exposures/require multiple DRS for multiple outcomes	Single PS useful for multiple outcomes with attention to risk factors for all outcomes in PS development
Interest in effect measure modification	Disease risk is a natural scale for evaluation	Although less natural than the risk scale, a potentially principled summary scale
Balance disease risk across covariates	DRS is a natural scale for stratification/matching	Stratification/matching on PS may provide secondary balance
Balance treatment preference across covariates	Stratification/matching on DRS may provide secondary balance	PS is a natural scale for stratification/matching
Exclude (trim) subjects in one treatment group without comparable, alternatively treated comparators	Potentially valuable to exclude high- or low-risk subjects without comparators	Potentially valuable to exclude subjects in PS tails without comparators
Relevance of the *C*-statistic	A high *C*-statistic provides some evidence of good performance in discriminating subjects who will versus those who will not develop the outcome	A high *C*-statistic can indicate clearly different indications for use of the compared treatments with possibly substantial areas of nonoverlap in PS distributions

DOI: 10.1002/pds

disease risk score estimated in the unexposed subjects only.[33] However, as pointed out by Hansen, the validity of the disease risk score estimated in this way is sensitive to model form, especially the assumption of a uniform treatment effect across categories of disease risk. Even modest treatment effect heterogeneity can induce bias in the overall treatment effect with this approach. If the treatment groups differ substantially on important covariates (which is akin to a clear distinction of treatment groups by means of a propensity score), these concerns are enhanced. Furthermore, inclusion of the exposure effect in the estimation of the disease risk score limits its value as a balancing score, as also discussed by Hansen.

Estimation of the disease risk score among unexposed subjects in the study population is also readily implementable, makes fewer assumptions than standard approaches that include exposed subjects, and yields a balancing score with desirable theoretical properties. However, reliable estimation of the model is a particular challenge in settings with relatively few outcomes, and these are expected among the unexposed subjects in the early monitoring period for a new therapy. Furthermore, if the disease risk score is used to form strata for estimation of treatment effects within levels of disease risk, the overfitting of the model under this approach tends to overestimate treatment benefits in the high-risk group and underestimate treatment harms in the low-risk group, which substantially limits the value of the score as an axis upon which to evaluate potential effect measure modification.

Furthermore, if the estimated disease risk score is strongly correlated with exposure status, the biases found by Pike and colleagues[34] to be associated with stratification by the disease risk score estimated by the approach of Miettinen also apply to disease risk scores estimated in the unexposed.[26] These concerns have probably contributed to the relatively infrequent use of disease risk scores in pharmacoepidemiology.[35] However, if exposed and unexposed subjects differ substantially on important determinants of disease risk, such that the shared support of risk factor distributions is limited, then valid comparison of treatments in an observational setting becomes less feasible,[36,37] and stratification on either a risk or propensity score is a useful way to identify such nonoverlap. Rather than a limitation, the ready ability to identify the kinds of subjects who almost always receive one specific treatment, and who thus should probably not be included in an analysis of comparative effectiveness, is a strength of both propensity score and disease risk score methods in pharmacoepidemiology.[35,36,38]

The disease risk score can also be estimated with data from a period prior to the study period or from a separate population. However, one difficulty with estimation in a separate population is that covariate assessments may differ from those in the study population. In the context of evaluation of a new therapy, the period before its introduction in the target study population may be useful. The reasoning behind this approach is that in times of evolving therapies, the disease risk in the population may be more stable than the propensity score. We illustrate this approach in the examples that follow.

EXAMPLE STUDY: INTRODUCTION OF A NEW STATIN OR MORE INTENSE STATIN THERAPY

We used data on statin therapy in patients after myocardial infarction to illustrate the development of a disease risk score in the context of new and evolving therapies. Large-scale randomized trials conducted between 1994 and 1998 demonstrated the value of statin therapy after myocardial infarction for prevention of recurrent myocardial infarction, stroke, and cardiovascular death.[39–41] Later trials showed that higher statin doses yielded greater risk reductions in this population.[42,43] We considered use of a disease risk score to evaluate the relative effectiveness of atorvastatin (Lipitor, the first high-intensity statin marketed in the USA), beginning with its first availability at the beginning of 1997; we also considered the efficacy of more intense statin therapy (defined according to the algorithm of Choudhry *et al.*[44]), which also was seldom used prior to this time.

We studied 5668 enrollees aged 65–100 years in either New Jersey's or Pennsylvania's state-sponsored pharmacy assistance program who survived a myocardial infarction and filled a statin prescription within 30 days after discharge between 1 January 1995 and 31 December 2004.[45,46] Figure 1 shows the strong time trend in the percentage of such first post-myocardial infarction prescriptions that were either atorvastatin or a high-dose statin. The efficacy endpoint was the composite including recurrent myocardial infarction, stroke, or death within 1 year after statin initiation. The analytic challenge was to develop an approach to control for multiple potential confounding variables that was applicable even during the early years of use of atorvastatin and high-dose statin therapy (i.e., 1997 and 1998) and consistent with confounder control in later experience. The study was approved by the institutional review board of Partners Healthcare.

DOI: 10.1002/pds

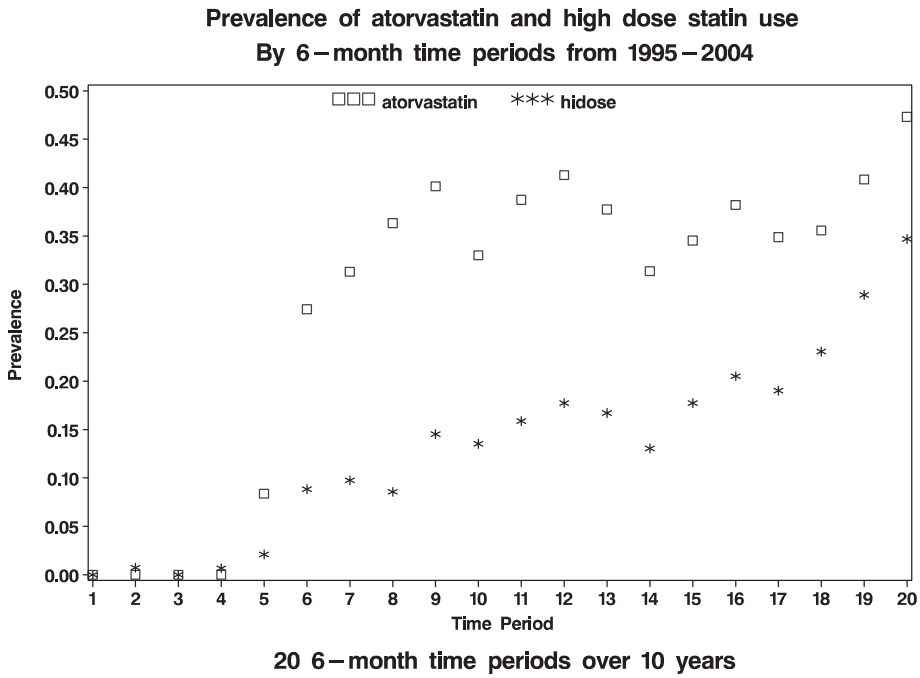

Figure 1. Prevalence of atorvastatin and high-dose statin use (by 6-month periods from 1995–2004)

The principles discussed earlier suggested development of a disease risk score based on the unexposed individuals in the period just before exposure availability. Thus, we used data from individuals who used statins other than atorvastatin or high-dose statins after myocardial infarction in 1995 and 1996, we but also included the individuals with exposure to these drugs following index myocardial infarction in 1997. We felt that the indications for exposure were still evolving and uncertain in that year, and the additional data would improve the reliability of the risk score. We used a logistic regression model to develop a disease risk score based on 826 patients who initiated statin therapy other than atorvastatin or high-dose statin, among whom 203 had recurrent myocardial infarction, stroke, or death within a year after statin initiation. Variables included in the disease risk score model were demographic characteristics, indicators of specific diseases encoded in medical encounters during the preceding year, summary measures of comorbidity (Charlson index and numbers of different generic drugs with prescription filled in the past year), and circumstances of the index hospitalization including angiography and duration of stay.

Table 2 shows the 20 variables included in the disease risk score; their prevalence or median in the population used to develop the score; and their multivariable relationship with recurrent myocardial infarction, stroke, or death. Variables associated with increased risk of the composite outcome were older age, Black race, history of heart failure, and higher Charlson score, whereas angiography during the index hospitalization and a diagnosis of hypertension were associated with reduced risk. Overall, the disease risk

Table 2. Baseline characteristics and contribution to the risk score: 826 initiators of a low-dose, non-atorvastatin statin post-MI from 1995 until 1997, 203 of whom had recurrent MI or stroke or died within 1 year

Variable	%	Odds ratio	95%CI
Age, per year, median (IQR)	76 (71–80)	1.05	1.02–1.08
Male	23.7	1.2	0.8–1.8
Black race	5.8	3.1	1.6–6.1
Other race	1.3	1.0	0.2–5.2
New Jersey resident	32.8	1.1	0.8–1.6
Congestive heart failure	54.0	1.7	1.1–2.4
Peripheral vascular disease	26.0	1.3	0.9–1.9
Chronic kidney disease	14.8	1.1	0.7–1.7
Hypertension	74.9	0.7	0.5–1.0
Diabetes	47.5	0.8	0.5–1.1
Prior myocardial infarction	7.5	0.9	0.5–1.7
Cerebrovascular disease	29.1	0.9	0.6–1.3
Prior hospital days, per day, median (IQR)	0 (0–3)	1.00	0.98–1.03
Charlson score, per point, median (IQR)	2 (1–3)	1.2	1.1–1.4
8–13 different medications	42.9	0.9	0.6–1.3
≥14 different medications	19.5	1.5	0.9–2.5
MI hospitalization 5–6 days	27.4	1.2	0.7–2.2
MI hospitalization 7–9 days	28.0	1.4	0.8–2.6
MI hospitalization ≥10 days	30.1	1.5	0.8–2.7
Angiography	55.3	0.6	0.4–0.8

Note. IQR, interquartile range; MI, myocardial infarction.

DOI: 10.1002/pds

score model had a *C*-statistic of 0.71 to predict 1-year risk of recurrent myocardial infarction, stroke, or death from any cause.

Based on this model, the predicted disease risk in statin initiators from 1997 through 2005 had a wide range, with mean predicted probability of 0.272 in atorvastatin initiators, 0.277 in initiators of other statins, 0.272 in initiators of high-dose statins, and 0.276 in initiators of lower-dose statins (Figures 2 and 3). Distributions of disease risk scores overlapped broadly and were similarly shaped across treatment groups. The slightly lower mean disease risk scores in the atorvastatin and high-dose statin groups reflected younger average ages, higher rates of angiography, and decreased prevalence of an index hospitalization lasting 10 or more days (Table 3).

EXAMPLE RESULTS

Several approaches are possible in the use of disease risk scores for confounder control in comparative effectiveness research, including matching exposure groups on risk levels, stratified analysis, and multivariate adjustment. Table 4 shows the impact of adjustment for the estimated disease risk score in comparisons of atorvastatin versus other statin regimens and of high versus lower doses of statins. Adjustment for disease risk as a continuous variable led to slight changes of crude estimates of 7%–8% reductions in the odds of recurrent myocardial infarction, stroke, or death associated with atorvastatin treatment or treatment with high-dose statins.

The disease risk score may have particular utility for the control of confounding in early follow-up after introduction of a new therapy. Parallel logistic regression analyses controlling for disease risk score and restricted to the 897 individuals who initiated statin therapy after myocardial infarction during the 2-year period 1997–1998 found that users of atorvastatin had somewhat lower risk relative to users of other statins (odds ratio, 0.71; 95%CI, 0.5–1.0) and that users of high-dose statins had reduced risk relative to users of lower-dose statin therapy (odds ratio, 0.57; 95% CI, 0.3–1.1).

Stratification on the disease risk score provides a straightforward approach to evaluate possible effect measure modification across levels of disease risk (Table 5). Observed risk of the composite outcome ranged across quintiles of the disease score from 12.6% to 31.4% in atorvastatin-treated patients and from 11.8% to 32.4% in patients treated with high-dose statins. Generally, odds ratios associated with atorvastatin therapy as well as with high-dose statin therapy tended toward greater reductions in higher-risk individuals, although confidence intervals were wide and broadly overlapping.

We also considered development of propensity scores as a strategy to control confounding in the

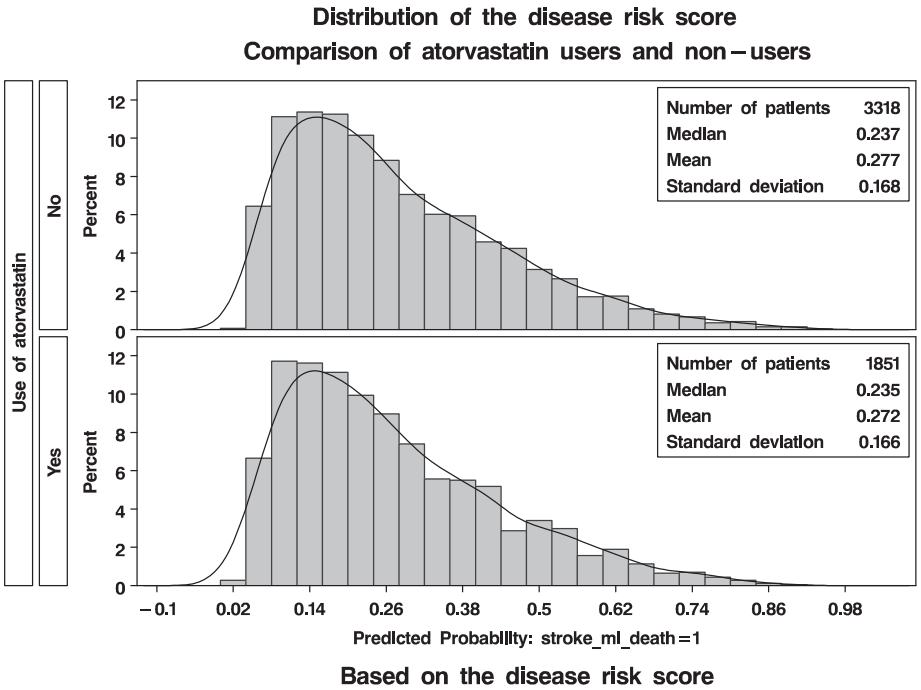

Figure 2. Distribution of the disease risk score. Comparison of atorvastatin users and nonusers

DOI: 10.1002/pds

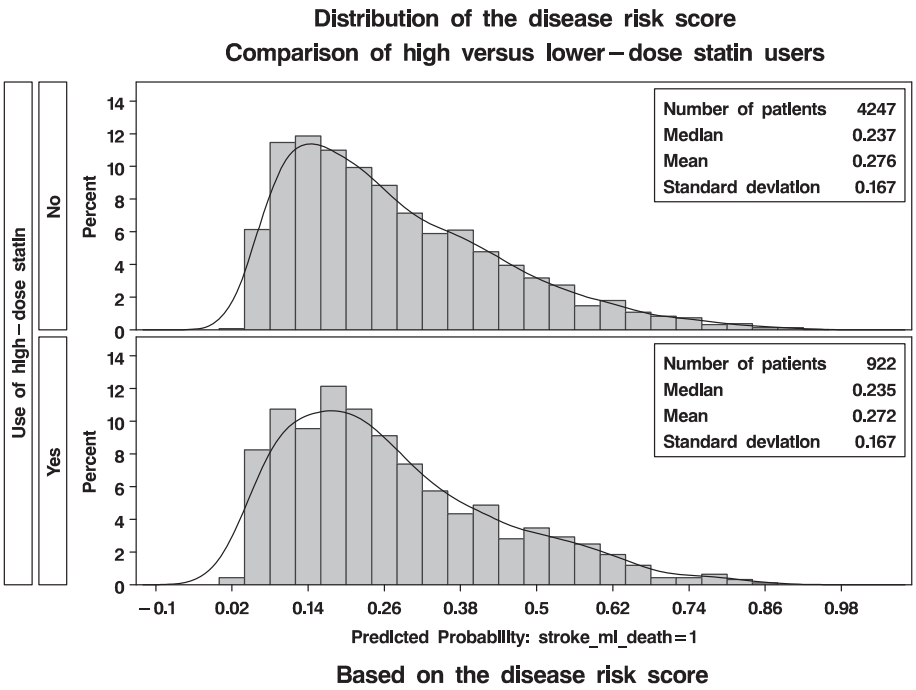

Figure 3. Distribution of the disease risk score. Comparison of high- versus lower-dose statin users

evaluation of relationships of atorvastatin and high-dose statin therapy with outcomes in these data. Challenges to the use of propensity scores in this setting included uncertainty about evolving prescriber preferences in the face of new evidence on benefits of these therapies during the study period and small numbers of individuals initiating high-dose statin therapy during early follow-up (Figure 1), which limited the ability to estimate period-specific propensity scores with all covariates included. In particular, with only 68 initiators of high-dose statins during the years 1997–1998, a logistic model predicting this treatment and including all 20 covariates showed evidence of overfitting and had coefficients with large differences from the estimates obtained in a model based on data from 1999–2004. In light of the small number of initiators of high-dose statins during early follow-up, it was unclear whether apparent differences from propensity score estimates in the later follow-up represented sampling variability or true changes in treatment preferences.

DISCUSSION

In comparative effectiveness and safety analyses with evolving therapies, the disease risk score may be a valuable tool to balance important covariates across treatment groups, to identify types of subjects with

Table 3. Comparison of baseline characteristics related to disease risk in 5169 statin initiators post-myocardial infarction between 1997 and 2005

	Atorvastatin	Non-atorvastatin	High-dose statin	Lower-dose statin
Variable, *n*	1851	3318	922	4247
Age, years, mean (SD)	78.2 (6.6)	78.7 (6.5)	77.7 (6.5)	78.7 (6.5)
Male, %	24.3	26.4	24.8	25.8
Black race, %	5.6	5.8	6.5	5.5
Angiography, %	62.9	61.2	63.9	61.3
Hypertension, %	80.4	81.0	83.1	80.3
Congestive heart failure, %	59.9	59.9	61.4	59.5
Peripheral vascular disease, %	27.1	27.3	28.7	26.9
Diabetes, %	51.5	49.1	55.3	48.8
Charlson score, mean (SD)	2.6 (2.0)	2.5 (2.0)	2.7 (2.1)	2.5 (2.0)
Myocardial infarction hospitalization ≥10 days, %	22.5	25.4	21.4	25.1
≥14 different medications, %	29.9	28.3	32.0	28.2
Disease risk score, mean (SD)	27.2 (16.6)	27.7 (16.8)	27.2 (16.7)	27.6 (16.7)

SD, standard deviation.

DOI: 10.1002/pds

Table 4. Crude and adjusted relative odds of recurrent myocardial infarction, stroke, or death within 1 year after initiation of statins among myocardial infarction survivors, 1997–2005; 5189 statin initiators, 1851 with atorvastatin, and 922 with high-dose statins

	Odds ratio	95%CI
Model: atorvastatin versus other		
Crude estimate	0.92	0.80–1.05
Adjusted for disease risk	0.93	0.81–1.07
Model: high dose versus other		
Crude estimate	0.93	0.78–1.11
Adjusted for disease risk	0.94	0.79–1.12

nonoverlap between treatment groups where valid comparisons of comparative effects are impossible, and to evaluate potential treatment effect measure modification.

Compared with propensity scores, disease risk scores are far less commonly used and have more theoretical shortcomings for comparative effectiveness research. In particular, although both approaches share the useful ability to reduce to one the dimension of potential confounders, balance with respect to the disease risk score can be evaluated only in the untreated, and estimation of this score within the study population is potentially problematic. Nonetheless, in the setting of early evaluation of evolving therapies, where reduction of the dimension of the confounder space is particularly desirable, no coherent propensity score may exist because of changing patient and provider preferences. The disease risk score is likely to be more stable over time, and the work required to estimate this score based on recent history in the health system under study may improve estimates of comparative effectiveness. Furthermore, although the disease risk score is less useful in settings with rare outcomes where reliable multivariable risk prediction is problematic, the risk score approach has advantages in studies of multiple exposures such as our consideration of atorvastatin and high-dose statin therapy, where a single score is applicable to all exposure categories.[47]

Another advantage of the disease risk score is its utility as a scale for arguably the most important dimension in the evaluation of possible effect measure modification. Absolute disease risk plays a critical role in many treatment decisions. Stratification on a disease risk score provides a transparent approach to compare absolute and relative treatment effects on this important axis.

However, one need not choose between a disease risk score and a propensity score approach to balance potential confounders. As an approach to match subjects in alternative treatment groups, one can minimize the distance in both the disease risk and propensity score dimensions. Neither must one weight the distances in these two dimensions equally. With common exposures and fewer data on outcomes, one can emphasize the distance on the propensity score dimension via overweighting. Conversely, with new or rare treatments, one can emphasize the disease risk score dimension.

In summary, we believe that the disease risk score is a useful tool with unique strengths for comparative safety and effectiveness research on new and evolving therapies. Evidence on comparative effectiveness of medications is particularly needed shortly after market approval, when insurance coverage decisions must be made. Products marketed with evidence of superior benefits or more favorable safety profiles, as compared with existing alternatives, will likely receive positive coverage conditions and therefore experience rapid uptake in the marketplace. Insurers seek timely comparative data to avoid fast and diffuse adoption of less effective or possibly harmful drugs; once prescribing patterns are established, they are difficult to change, even in the face of compelling comparative effectiveness evidence.

Table 5. Relationship of atorvastatin use and use of high-dose statins with risk of recurrent myocardial infarction, stroke, or death within 1 year, stratified by quintiles of the disease risk score

Predicted risk	Observed	Atorvastatin		Other statin		Odds ratio (95%CI)
	Events/*n*	*n*, %	Events, risk	*n*, %	Events, risk	
3.5–12.7	116/1033	381, 36.9	48, 12.6	652, 63.1	68, 10.4	1.21 (0.9–1.7)
12.7–19.8	176/1034	375, 36.3	59, 15.7	659, 63.7	117, 17.8	0.89 (0.7–1.2)
19.8–28.3	216/1034	369, 35.7	71, 19.2	665, 64.3	145, 21.8	0.88 (0.7–1.1)
28.3–41.3	292/1034	366, 35.4	108, 29.5	668, 64.6	184, 27.5	1.07 (0.9–1.3)
41.4–94.0	363/1034	360, 34.8	113, 31.4	674, 65.2	250, 37.1	0.85 (0.7–1.0)
		High-dose statin		Low-dose statin		
3.5–12.7	116/1033	195, 18.9	23, 11.8	838, 81.1	93, 11.1	1.06 (0.7–1.6)
12.7–19.8	176/1034	180, 17.4	36, 20.0	854, 82.6	140, 16.4	1.22 (0.9–1.7)
19.8–28.3	216/1034	191, 18.5	36, 18.9	843, 81.5	180, 21.4	0.88 (0.6–1.2)
28.3–41.3	292/1034	177, 17.1	45, 25.4	857, 82.9	247, 28.8	0.88 (0.7–1.2)
41.4–94.0	363/1034	179, 17.3	58, 32.4	855, 82.7	305, 35.7	0.91 (0.7–1.1)

DOI: 10.1002/pds

Initiatives such as the Sentinel System of the Food and Drug Administration reflect a heightened interest in early identification of adverse effects and benefits of drugs and new doses as soon after marketing as possible. Joint consideration of both propensity and disease risk scores for new therapeutics has the potential to improve estimates of comparative effectiveness.

CONFLICT OF INTEREST

Dr Glynn has received grants from AstraZeneca and Novartis for clinical trials monitoring and analysis and has been invited to Grand Rounds at Merck.

Dr Schneeweiss is a principal investigator of the Brigham and Women's Hospital DEcIDE Center on Comparative Effectiveness Research and the DEcIDE Methods Center (both funded by AHRQ) and of the Harvard-Brigham Drug Safety and Risk Management Research Center (funded by the FDA). Dr Schneeweiss is a paid member of the Scientific Advisory Board of HealthCore and consultant to WHISCON and Booz & Co.; in addition, he is the recipient of investigator-initiated grants from Pfizer and Novartis.

KEY POINTS

- Use of propensity scores or multivariable models to control confounding in the evaluation of risks and benefits of new or evolving therapies can be problematic due to evolving indications and small numbers of outcomes in early follow-up.
- Availability of data on confounders and the study outcome before introduction of the new therapy allows for development of a disease risk score with important balancing properties.
- The disease risk score can help identify the range of alternatively treated subjects in whom evaluation of comparative effectiveness and safety are possible, and provides an important dimension for evaluation of effect measure modification.

ACKNOWLEDGEMENTS

This study was supported by Grants AG018833 and AG023178.

REFERENCES

1. Schneeweiss S, Gynn RJ, Avorn J, Solomon DH. A Medicare database review found that physician preferences increasingly outweigh patient characteristics as determinants of first-time prescriptions for COX-2 inhibitors. *J Clin Epidemiol* 2005; **58**: 98–102.
2. Harrell FE, Lee KL, Califf RM, Pryor DB, Rosati RA. Regression modeling strategies for improved prognostic prediction. *Stat Med* 1984; **3**: 143–152.
3. Rubin DB. Estimating causal effects from large data sets using propensity scores. *Ann Intern Med* 1997; **127**: 757–763.
4. Cepeda MS, Boston R, Farrar JT, Strom BL. Comparison of logistic regression versus propensity score when the number of events is low and there are multiple confounders. *Am J Epidemiol* 2003; **158**: 280–287.
5. Schneeweiss S, Gagne JJ, Glynn RJ, Ruhl M, Rassen JR. Assessing the comparative effectiveness of newly marketed medications: methodological challenges and implications for drug development. *Clin Pharmacol Ther* 2011; **90**: 777–790.
6. Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Stürmer T. Variable selection for propensity score models. *Am J Epidemiol* 2006; **163**: 1149–1156.
7. Myers JA, Rassen JA, Gagne JJ, *et al*. Effects of adjusting for instrumental variables on bias and precision in effect estimates. *Am J Epidemiol* 2011; **174**: 1213–1222.
8. Pearl J. Invited commentary: Understanding bias amplification. *Am J Epidemiol* 2011; **174**: 1223–1227.
9. Freedman AN, Yu B, Gail MH, *et al*. Benefit/risk assessment for breast cancer chemoprevention with raloxifene or tamoxifen for women age 50 years or older. *J Clin Oncol* 2011; **29**: 2327–2333.
10. Gail MH, Constantino JP, Bryant J, *et al*. Weighing the risks and benefits of tamoxifen treatment for preventing breast cancer. *J Natl Cancer Inst* 1999; **91**: 1829–1846.
11. National Cholesterol Education Program. ATP III Guidelines At-A-Glance Quick Desk Reference. US Department of Public Health, Public Health Service, National Institute of Health, National Heart, Lung, and Blood Institute. NIH Publication No. 01–3305. Accessed July 11, 2011 at www.nhlbi.nih.gov/guidelines/cholesterol/atglance.pdf.
12. Knaus WA, Wagner DP, Draper EA, *et al*. The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults. *Chest* 1991; **100**: 1619–1636.
13. Lyden P, Lu M, Jackson C, *et al*. Underlying structure of the National Institutes of Health Stroke Scale: results of a factor analysis. NINDS tPA Stroke Trial Investigators. *Stroke* 1999; **30**: 2347–2354.
14. Teasdale G, Jennett B. Assessment of coma and impaired consciousness. A practical scale. *Lancet* 1974; **2**(7872): 81–84.
15. Ridker PM, Macfadyen JG, Nordestgaard BG, *et al*. Rosuvastatin for primary prevention among individuals with elevated high-sensitivity c-reactive protein and 5% to 10% and 10% to 20% 10-year risk. Implications of the Justification for Use of Statins in Prevention: an Intervention Trial Evaluating Rosuvastatin (JUPITER) trial for "intermediate risk". *Circ Cardiovasc Qual Outcomes* 2010; **3**: 447–452.
16. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* 1987; **40**(5): 373–383.
17. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. *Med Care* 1998; **36**: 8–27.
18. Schneeweiss S, Wang PS, Avorn J, Glynn RJ. Improved comorbidity adjustment for predicting mortality in Medicare populations. *Health Serv Res* 2003; **38**: 1103–1120.
19. Gagne JJ, Glynn RJ, Avorn J, Levin R, Schneeweiss S. A combined comorbidity score predicted mortality in elderly patients better than existing scores. *J Clin Epidemiol* 2011; **64**: 749–759.
20. Schneeweiss S, Rassen JA, Glynn RJ, Avorn J, Mogun H, Brookhart MA. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. *Epidemiology* 2009; **20**: 512–522.
21. Peters CC. A method of matching groups for experiments with no loss of populations. *J Ed Research* 1941; **34**: 606–612.
22. Belson WA. A technique for studying the effects of a television broadcast. *Applied Stat* 1956; **5**: 195–202.
23. Cochran WG. The use of covariance in observational studies. *J Roy Statist Soc. Ser C* (Applied Statistics). 1969; **18**: 270–275.
24. Graubard BI, Rao RS, Gastwirth JL. Using the Peters-Belson method to measure health care disparities from complex survey data. *Stat Med* 2005; **24**: 2659–2568.
25. Miettinen OS. Stratification by a multivariate confounder score. *Am J Epidemiol* 1976; **104**: 609–620.
26. Hansen BB. The prognostic analogue of the propensity score. *Biometrika* 2008; **95**: 481–488.
27. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983; **70**: 41–55.
28. Rosenbaum PR, Rubin DB. Reducing bias in observational studies using subclassification on the propensity score. *J Am Statist Assoc* 1984; **79**: 516–524.
29. Joffe MM, Rosenbaum PR. Invited commentary: propensity scores. *Am J Epidemiol* 1999; **150**: 327–333.
30. Cook EF, Goldman L. Performance of tests of significance based on stratification by a multivariate confounder score or by a propensity score. *J Clin Epidemiol* 1989; **42**: 317–324.
31. Stürmer T, Schneeweiss S, Brookhart MA, Rothman KJ, Avorn J, Glynn RJ. Analytic strategies to adjust confounding using exposure propensity scores and

disease risk scores: nonsteroidal anti-inflammatory drugs and short-term mortality. *Am J Epidemiol* 2005; **161**: 891–898.
32. Arbogast PG, Kaltenbach L, Ding H, Ray WA. Adjustment for multiple cardiovascular risk factors using a summary risk score. *Epidemiology* 2008; **19**: 30–37.
33. 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**: 613–620.
34. Pike MC, Anderson J, Day N. Some insights into Miettinen's multivariate confounder score approach to case–control study analysis. *Epidemiol Comm Health* 1979; **33**: 104–106.
35. Glynn RJ, Schneeweiss S, Stürmer T. Indications for propensity scores and review of their use in pharmacoepidemiology. *Basic Clin Pharm Toxicol.* 2006; **98**: 253–259.
36. Stürmer T, Rothman KJ, Avorn J, Glynn RJ. Treatment effects in the presence of unmeasured confounding: dealing with observations in the tails of the propensity score distribution--a simulation study. *Am J Epidemiol* 2010; **172**: 843–854.
37. Longmore RB, Yeh RW, Kennedy KF, *et al*. Clinical referral patterns for carotid artery stenting versus carotid endarterectomy: results from the Carotid Artery Revascularization and Endarterectomy Registry. *Circ Cardiovasc Interv* 2011; **4**: 88–94.
38. Schneeweiss S. A basic study design for expedited safety signal evaluation based on electronic healthcare data. *Pharmaceopidemiol Drug Saf* 2010; **19**: 858–868.
39. The Scandinavian Simvastatin Survival Study Group. Randomized trial of cholesterol lowering in 4444 patients with coronary heart disease: the Scandinavian Simvastatin Survival Study (4S). *Lancet* 1994; **344**: 1383–1389.
40. Sacks FM, Pfeffer MA, Moye LA, *et al*. Cholesterol and Recurrent Events Trial Investigators. The effect of pravastatin on coronary events after myocardial infarction in patients with average cholesterol levels. *N Engl J Med* 1996; **355**: 1001–1009.
41. The LIPID Study Group. Prevention of cardiovascular events and death with pravastatin in patients with coronary heart disease and a broad range of initial cholesterol levels. *N Engl J Med*. 1998; **339**: 1349–1357.
42. Cannon CP, Braunwald E, McCabe CH, *et al*. Intensive versus moderate lipid lowering with statins after acute coronary syndromes. *N Engl J Med* 2004; **350**: 1495–1504.
43. LaRosa JC, Grundy SM, Waters DD, *et al*. Intensive lipid lowering with atorvastatin in patients with stable coronary disease. *N Engl J Med* 2005; **352**: 1425–1435.
44. Choudhry NK, Levin R, Winkelmayer WC. Statins in elderly patients with acute coronary syndrome: an analysis of dose and class effects in typical practice. *Heart* 2007; **93**: 945–951.
45. Setoguchi S, Glynn RJ, Avorn J, Levin R, Winkelmayer WC. Ten-year trends of cardiovascular drug use after myocardial infarction among community-dwelling persons > or =65 years of age. *Am J Cardiol* 2007; **100**: 1061–1067.
46. Setoguchi S, Glynn RJ, Avorn J, Mittleman MA, Levin R, Winkelmayer WC. Improvements in long-term mortality after myocardial infarction and increased use of cardiovascular drugs after discharge: a 10-year trend analysis. *J Am Coll Cardiol* 2008; **51**: 1247–1254.
47. Cadarette SM, Gagne JJ, Solomon DH, Katz JN, Stürmer T. Confounder summary scores when comparing the effects of multiple drug exposures. *Pharmacoepidemiol Drug Saf* 2010; **19**: 2–9.