Assessing the Comparative Effectiveness of Newly Marketed Medications: Methodological Challenges and Implications for Drug Development

S Schneeweiss[1], JJ Gagne[1], RJ Glynn[1], M Ruhl[2] and JA Rassen[1]

Comparative-effectiveness research (CER) aims to produce actionable evidence regarding the effectiveness and safety of medical products and interventions as they are used outside of controlled research settings. Although CER evidence regarding medications is particularly needed shortly after market approval, key methodological challenges include (i) potential bias due to channeling of patients to the newly marketed medication because of various patient-, physician-, and system-related factors; (ii) rapid changes in the characteristics of the user population during the early phase of marketing; and (iii) lack of timely data and the often small number of users in the first few months of marketing. We propose a mix of approaches to generate comparative-effectiveness data in the early marketing period, including sequential cohort monitoring with secondary health-care data and propensity score (PS) balancing, as well as extended follow-up of phase III and phase IV trials, indirect comparisons of placebo-controlled trials, and modeling and simulation of virtual trials.

Comparative-effectiveness research (CER) aims to produce actionable evidence regarding the effectiveness and safety of medical products and interventions as they are used outside of controlled research settings (**Figure 1a**).[1] The ultimate goal is to support optimal decision-making by stakeholders in the health-care system, including patients, physicians, provider organizations, industry, and insurers. Although ~50% of drugs newly approved by the US Food and Drug Administration (FDA), including anti-infectious medications and anti-neoplastic agents, undergo some sort of active comparator analysis in preapproval trials,[2] such information is often insufficient to answer all questions regarding optimal prescribing of these new agents. CER—using secondary health-care data, including electronic medical records (EMRs), longitudinal claims data, and registries—offers the benefit of studying outcomes of these medicines under the conditions of routine medical practice without intervening in the delivery of the care that gives rise to the data. However, owing to inherent methodological limitations in this approach, additional research strategies will need to be applied, including randomized trials, in which greater validity is usually traded off against generalizability of findings to the day-to-day practice of medicine.

With newly marketed agents, the sooner valid CER results can be produced, the more useful they are to all stakeholders (**Figure 1b**). Insurance coverage decisions must be made shortly after marketing authorization. Products marketed with evidence of benefits and/or tolerability superior to existing alternatives will likely receive preferential coverage in health plans (e.g., reduced copayment for patients) and therefore will be taken up more rapidly by the marketplace. On the other hand, 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 evidence of comparative effectiveness (CE). For example, although ezetimibe has not been proven superior to statins,[3] it has been used widely since its approval.[4] Many countries already make coverage decisions based on evidence of incremental CE,[5] and in the United States the FDA and the Center for Medicare and Medicaid Services have agreed to work together more closely to allow for simultaneous review of drug approval and coverage.[6] Consequently, generation and synthesis of CE information will become increasingly important to portfolio management early in the development process[7] and

[1]Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA; [2]Booz & Co, Düsseldorf, Germany. Correspondence: S Schneeweiss (schneeweiss@post.harvard.edu)

Received 8 August 2011; accepted 30 August 2011; advance online publication 2 November 2011. doi:10.1038/clpt.2011.235

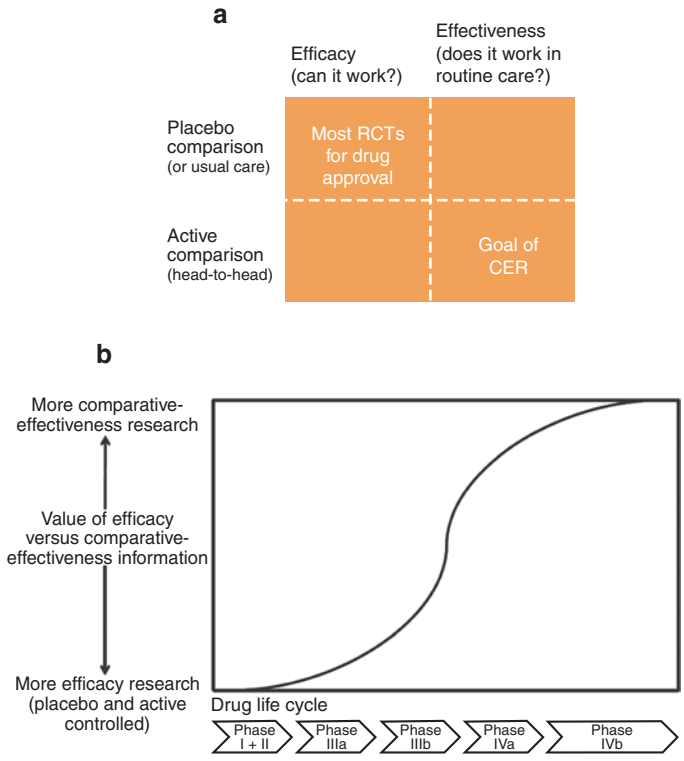

Figure 1 Comparative-effectiveness research (CER). (**a**) Goal of CER in contrast to preapproval randomized controlled trials (RCTs). (**b**) The increasing value of CER during the drug life cycle.

will ultimately affect the financial valuations of pharmaceutical companies.[8]

Observational (nonrandomized) studies fill a critical gap in the CER landscape.[9] However, these studies face many challenges,[10] which are amplified when such studies are used to compare the effectiveness of newly marketed medications. Key methodological challenges include (i) potential bias due to channeling of patients to the newly marketed medication because of patient-, provider-, and system-related factors; (ii) a quickly shifting user population with varying levels of background risk during the early phase of marketing;[11] and (iii) the lack of timely data and the often small number of users in the first few months of marketing, the latter of which reduces the precision of effect estimation and limits subgroup analyses. Of these challenges, channeling is often the biggest threat to the validity of nonrandomized studies and may become more extreme if the medication of interest is a first-in-class agent.

Given that a drug's value to society is critically influenced by its performance outside of a controlled research environment, decisions on how CER evidence will be generated must be made as early as possible. In this article, we characterize the methodological challenges faced when assessing the CE of newly marketed medications as they are used in day-to-day care. We focus our discussion on what we call phase IVA activities, those occurring during the transitional time between a product's marketing authorization and the achievement of the "quiescent state" in which use patterns, market share, and insurance coverage have stabilized. This transitional period may last 6 months after marketing for some drugs and 2–3 years for others, depending on the speed of uptake, the generation of actionable evidence, and other factors.

In light of these challenges, we propose a framework that integrates evidence from multiple sources for assessing the CE of drugs in the early marketing period. This is followed by a discussion of the opportunities and challenges for a drug development process whose goal is to establish the CE of products early during marketing; we also make suggestions for structural changes throughout the development process in order to support early generation of CE evidence.

METHODOLOGICAL CHALLENGES

Owing to the absence of CER results in the early marketing setting—when multiple stakeholders seek such data to inform decision-making—payers increasingly conduct their own postmarketing observational CE studies based on longitudinal claims data from their own enrollee population.[12] However, nonrandomized research on the effectiveness of medications is methodologically challenging, and conducting such studies in a new-to-market setting only increases the difficulty. In this section, we consider the general challenges of nonrandomized CER as well as challenges that are specific to newly marketed medications.

Selective prescribing of drugs that are new to the market will lead to confounding

Where a patient is well controlled on an existing medication, there is little incentive to change the current treatment. On the other hand, new medications create expectations of improved

effectiveness and tolerability, particularly among patients for whom existing therapies may not have performed optimally. Consequently, early users of a newly marketed medication may not be representative of a drug's eventual user population.[13] Any imbalances in disease severity, prognosis, or risk profile between users of the new drug and comparison patients may bias effect estimates in nonrandomized studies, unless such confounding factors can be fully adjusted for or, if specific, homogeneous patient subgroups can be identified.[14] In this section, we address the effects of selective prescribing on study validity and then suggest methods to combat the resulting potential for bias.

As health-care professionals, we like to think that prescribing decisions are precisely tailored to each patient's expected benefits and risks. If this were true, however, we would have a fully deterministic model for treatment choice; if this model were codified in a universally accepted and strictly followed treatment guideline, there would remain no random variation in patients' treatment assignment. In this scenario, all treatment–outcome associations would be intractably confounded and nonrandomized CER would be impossible. On the other hand, CER would be unnecessary, because any relevant knowledge about the relative performance of medications would have already been incorporated into the treatment choice model.

Fortunately for epidemiologists (but to the frustration of guideline writers), this is rarely the case, and we often observe substantial variations in treatment assignment that cannot be explained by patient characteristics. Epidemiology seeks to identify and exploit this random variation in exposure status by comparing patients with concordant baseline risk but discordant exposure status for the outcome of interest.[1] The assumption underlying successful nonrandomized CER is that we can measure and assess patients' baseline risk in order to identify valid differences between treatment groups.

Health-care systems and the care decisions made within them are inherently hierarchical. Patient-, provider-, and system-level factors can all affect whether particular patients receive new drugs (**Figure 2**). If these factors contribute to variation in exposure and are also causes of an outcome of interest, then they are confounders that need to be addressed. If these factors contribute to variation in exposure but are unrelated to the outcome, then they contribute to exposure variation that can be harnessed and exploited using instrumental variable estimation in an effectiveness study. However, determining the category into which specific factors fall is often difficult.[10]

Patient-level confounding. Patient characteristics that drive new drug use decisions can vary from drug to drug, and the early users of newly marketed medications may be selective with respect to factors related to both expected effectiveness and tolerability of the new medication.

In medications such as statins, of which there are many agents in a class, the first patients to use a new agent are likely to be those who have suffered side effects from existing agents or who have not achieved sufficient low-density lipoprotein cholesterol control. The first patients with rheumatoid arthritis to use a new immunomodulating drug are probably those who experienced little or no benefit from existing drugs and may therefore respond to the new drug in a way that the average patient would not. Patients with atrial fibrillation who are using a new direct thrombin inhibitor are likely to be those in whom coagulation could not be well controlled with warfarin or patients who find the monitoring requirements or side effects of warfarin intolerable.

When exenatide (Byetta) came to market, Segal *et al.* found that users of the drug during its first 3 months after marketing had a history of more visits to a physician in the previous year, had a slightly higher level of glycolated hemoglobin, and had used insulin and oral antidiabetic medications more often, as compared to patients who were initiated 6 or 7 months after marketing.[13] We observed similar selective prescribing to sicker patients in the first quarter of marketing of selective cyclooxygenase-2 inhibitors, which are analgesics targeted to patients with preexisting upper-GI complaints.[11] It may not be feasible to find a suitable comparison group for a nonrandomized study of patients in whom treatment with a new direct thrombin inhibitor is initiated, given that the population of first users of the new drug will likely be enriched with those who have previously found warfarin therapy to be problematic. Therefore, early users of newly marketed medications may be highly differential with respect to the benefits and adverse effects of alternative treatments.

In the extreme case, patient populations are simply not comparable in the immediate postmarketing period. If, after estimation

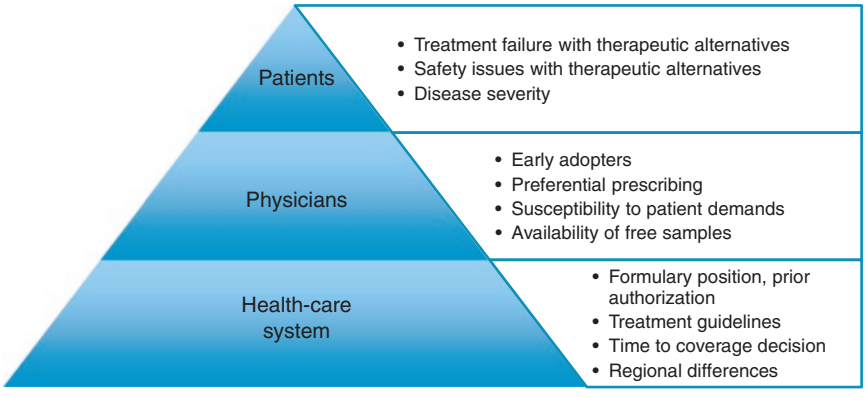

Figure 2 The multilevel nature of factors that determine prescribing of newly marketed medications.

of a propensity score (PS)—the predicted probability of receiving the new drug over an old treatment, conditional on all observed patient characteristics[15]—a plot of the PS distributions shows absence of overlap (**Figure 3**), we can infer that for no patient who received the new drug would the old drug have been a reasonable treatment choice, and vice versa. In other words, there is no "equipoise" between the two agents: because some characteristic or set of characteristics completely determined the patients' treatment choice, these patients would never have been deemed comparable in a clinical trial. These characteristics may come from areas beyond traditional exclusion criteria for randomized clinical trials (RCTs); they could be patient-related (e.g., nonresponse to earlier therapies), physician-related (no experience with the new drug), or system-related (no coverage for the new drug).

As time goes on, such selective use may diminish as a result of actual or perceived evidence of benefit of the new drug in the broader population. General uptake of the drug will make patients receiving the new drug more similar to those on the older therapy, and the PS distributions will migrate toward each other, increasing the amount of overlap (**Figure 4**). With substantial overlap, a larger number of comparable patients can be identified. We must assume that all confounders are observed; without this assumption, the large overlap in distributions may be an artifact of omitting an important discriminating factor from the PS model.[10]

Validity/generalizability trade-off: Unlike in most clinical trials, the goal of CER is to be generalizable to the broad group of patients receiving therapies outside of controlled research environments.[16] However, in the early marketing phase, the scenario represented in **Figure 4** may be common. Initially there may be only a small subgroup of patients—those whose PSs overlap—who are similar with respect to a wide range of characteristics and can be validly compared. These patients may sometimes be easily identifiable (e.g., they may be of a particular age group); in other cases, the composition of the subgroup may be more abstract (e.g., patients who did not respond well to earlier therapies). Whatever the case, it may be necessary to impose stringent restrictions before attempting to make a valid inference from the data, thereby inevitably limiting the generalizability of the CE results. Indeed, these restrictions may yield patient populations that resemble those studied in RCTs. It has been repeatedly demonstrated that similar restrictions in nonrandomized CER studies resulted in findings comparable to those of RCTs.[14,17]

More broadly, as Psaty notes, the validity/generalizability trade-off is a familiar phenomenon but should nonetheless be considered when we speak about CER and its goal of being as generalizable as possible to all patients in routine care.[18] Although we support that goal, we ultimately choose validity over generalizability—not as an academic exercise but as a matter of good practice—when making treatment choices and public health recommendations. The cost of the trade-off may not be as great as it is often perceived because, from a clinical perspective, the most important generalization to make is about the comparability of therapies among patients for whom either of the drugs would be a reasonable treatment choice. CER among these patients may not yield the most generalizable information, but it will yield information that is most relevant to clinical decision making.

Effect measure modification: So far we have interpreted the nonoverlap in **Figure 4** as being driven purely by confounding factors. As the user population shifts over time toward the "average" patient, we generally expect the influence of confounding to diminish. However, changes in the composition of the user population over time have ramifications not only for confounding but also for the introduction of possible effect-measure modification. As the composition of the user group

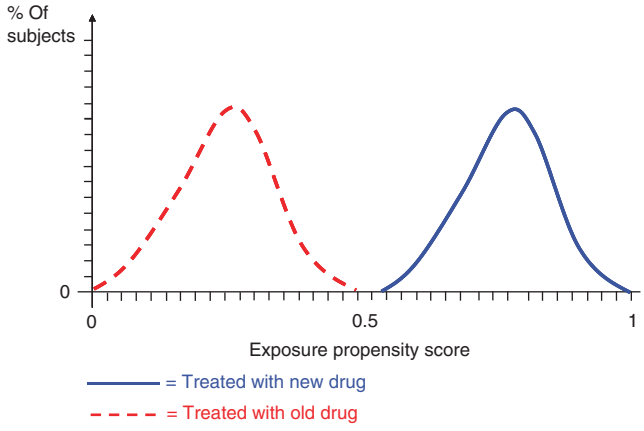

Figure 3 Illustration of noncomparable patient populations. Each patient—whether treated with the new drug or the old drug—has an estimated probability (or propensity) of treatment with the new drug that is conditional on all observed patient characteristics. This propensity ranges from 0 to 1 and is plotted separately for the two treatment groups. In this scenario, no patient in the new drug group has the same propensity score as any patient in the old drug group.

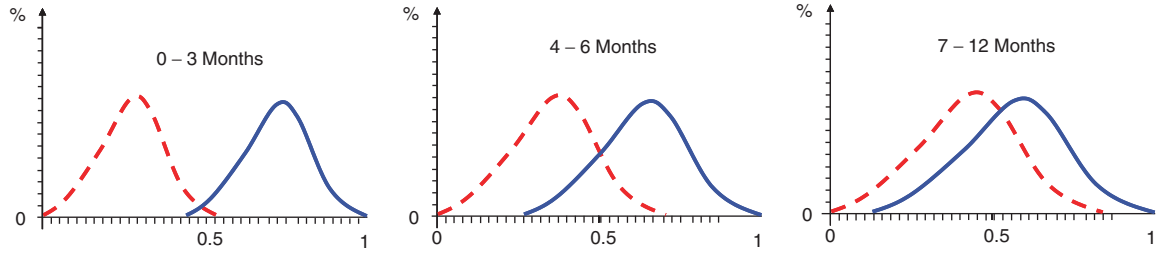

Figure 4 Increasing comparability of treatment groups with increasing market availability of a newly marketed drug. With increasing market penetration, the initiators of the new drug (blue solid curve) become more similar to those initiated into treatment with the older competitor drug (red broken curve), as indicated by increasingly overlapping propensity score distributions. The time intervals used here are for illustration purposes and may vary by drug and condition of interest.

changes, the baseline risk scores of the outcomes of interest (intended and unintended) may vary. In that event, a change in population-wide point estimates might reflect the fact that the drug truly has biologically different effects in different patients, and the observed difference is a causal modification of the treatment effect measure.[19] Although it is theoretically possible to identify such effect-measure modification in the early treatment period, in practice there may not be adequate numbers of treated subjects during this period for reliable subgroup estimation.

Physician-level confounding. Physician-level confounding stems from differences in the treatment choices two physicians may make for the same patient. With new drugs, some prescribers are early adopters; indeed, some physicians would almost automatically prescribe the newest medication when it becomes available. Other physicians may strongly prefer older medications with proven track records.

In a Medicare population (≥65 years of age) with insurance coverage through a pharmacy assistance program involving no restrictions or minimum copayments, we studied physicians' adoption of cyclooxygenase-2 inhibitors (coxibs) as an alternative to traditional nonselective nonsteroidal anti-inflammatory drugs (NSAIDs).[11] **Figure 5a** illustrates the fast adoption of coxibs among NSAID prescribers. Within two calendar quarters, half of the physicians had prescribed a coxib at least once, but even after two years, 20% of the prescribers had still never prescribed one. Conversely, we also identified physicians who always prescribed coxibs and never chose traditional nonselective NSAIDs (**Figure 5b**). One year after coxibs entered the market, 40% of the physicians fell into this group.

Although these variations in treatment may suggest confounding, it is important to distinguish between physician factors that influence only exposure vs. those that influence both exposure and the study outcome. The latter is true confounding and can be handled with matching, stratification, regression, or other standard techniques. The former case—when the factor predicts exposure but not outcome—is actually an example of an instrumental variable. These instrumental variables should not be treated as confounders; doing so in the presence of residual confounding may actually increase bias.[20,21] Instead, they can be exploited as a source of natural variation and used to obtain an unbiased effect estimate, even in the presence of unmeasured confounding.

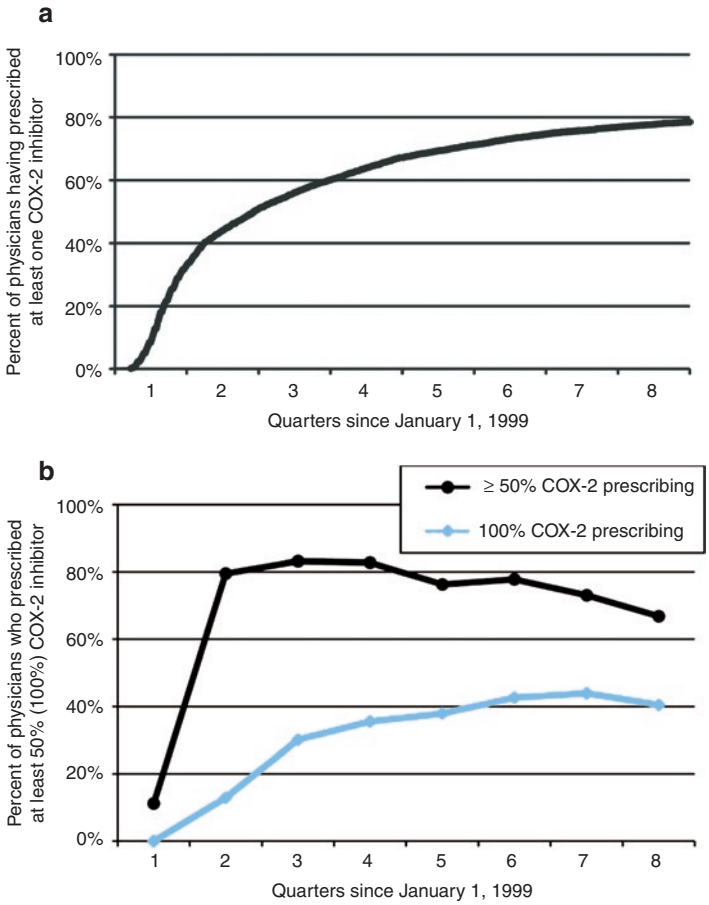

Figure 5 Rapid adoption of selective cyclooxygenase-2 inhibitors (coxibs) by prescribers as alternatives to traditional nonselective nonsteroidal anti-inflammatory drugs (NSAIDs) after coxibs entered the market in January 1999. (**a**) Time to first selective coxib prescription among all 6,972 prescribers of NSAIDs. (**b**) Proportion of prescribers who chose coxibs instead of traditional nonselective NSAIDs as the first NSAID prescription for their patients in 50% (100%) of instances (restricted to 464 high-volume prescribers with at least three first-time NSAID prescriptions per quarter).

As an example, we have previously argued that if physicians with a strong preference for coxibs are comparable to those who strongly prefer traditional nonselective NSAIDs in all other aspects that may influence patients' health outcomes, and that if patients select physicians without knowledge of their preference, then the prescribing preference of the physician is an instrumental variable for which standard methods can be used to obtain unbiased effect estimates.[1,22–24] On the other hand, if the differences between prescribers who prefer coxibs and those who prefer traditional nonselective NSAIDs are associated with disease severity (e.g., if sicker patients consult specialists who may be early adopters) or treatment outcomes (e.g., if early adopters provide better care in other ways that improve prognosis), then these physician-level characteristics should be treated as confounders.[25,26]

The prescribing preferences of physicians with respect to newly marketed medications may be further influenced by direct-to-consumer advertising.[27,28] To remain competitive, or simply to please patients, some physicians may be more likely to respond to patients' demand for newly marketed drugs. If physicians' response to direct-to-consumer advertising is differentially linked with patient characteristics and health outcomes, it may be a source of bias. We note that this phenomenon is hypothetical at this point, and, to our knowledge, it has not yet been observed in practice.

System-level confounding. All patients and physicians operate within health-care systems that impose additional influences/constraints on treatment decisions, affect the uptake of newly marketed medications, and may determine which patients receive newly marketed drugs. These factors may often be uncorrelated or only weakly correlated with patients' potential outcomes and therefore will not induce strong confounding. A major system-level determinant is restriction on insurance coverage and accessibility: an unfavorable formulary position may require high patient copayments, prior authorization, and step therapy, all of which limit access to coverage.[29] Indeed, in the early months, certain plans may simply not cover new-to-market drugs at all. Beyond the plan level, local or national provider communities produce treatment guidelines that may channel patients into preferred treatment options.

Sparse data during the early marketing phase

Unless a newly marketed medication has a spectacularly successful launch, it will take time for enough users to accrue to enable a direct comparison of a new drug with an old one. The causes of slow uptake are largely the same as the factors (described above) that can induce potential confounding: a rare condition/therapeutic target, a narrow indication, high drug cost, lack of payer coverage, limited perceived effectiveness, and an unfavorable safety profile. The early marketing phase (phase IVA) is dynamic, and some of these factors may quickly shift as additional CE information becomes available.

Several statistical aspects of CER are affected by slow uptake of a newly marketed medication. Most obviously, studies based on a small number of patients will yield imprecise effect estimates. The precision of estimates is further compromised if the outcome of interest is infrequent, such as reduction in the rate of myocardial infarction among patients without symptomatic coronary heart disease. Moreover, decision makers want not only to understand the average effectiveness of a new drug in a population but also to identify segments of the population in whom it works best and is tolerated well. Coverage may be extended to patients in whom benefits can be demonstrated, and these may constitute a subgroup of patients already covered for the older medications. Sparse data will limit the number of meaningful subgroup analyses possible; however, as use of a new drug becomes more frequent, it becomes increasingly feasible to investigate effects in more subgroups (**Figure 6**).

Not only does small study size reduce the precision of effect estimates, it may also hinder one's ability to control for confounding in a study involving the CE of drugs. It is well known that, for each variable included in a traditional multivariable regression outcome model, one must observe a certain number of outcomes in the study population.[30] As a remedy, PS methods enable the adjustment for a large number of confounders even if end points are rare.[31,32] In the setting of small samples (due to a large number of subgroups), the PS derived from the total population can be used to balance the subgroups if appropriate modeling strategies are implemented.[33]

Nevertheless, PSs have their limits. Confounding is likely to be strong in comparisons of newly marketed medications with established drugs, requiring adjustment of an even greater number of variables than PSs and other dimension-reduction techniques can accommodate in sparse-data settings. On the other hand, even if a reasonable PS model can be fitted, and

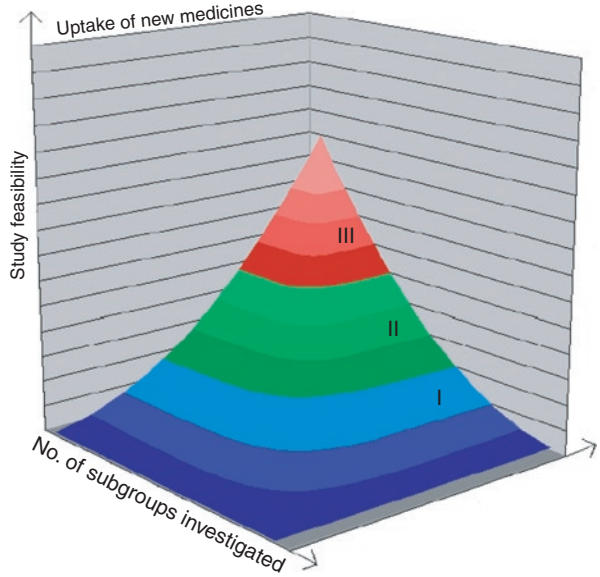

Figure 6 Feasibility of conducting postmarketing comparative-effectiveness research (CER) as a function of uptake and the number of subgroups of interest. In the early marketing phase, or if market uptake of a new drug is slow (blue areas, I), CER is challenging, and it is not feasible to investigate subgroups of patients regarding the drug's comparative effectiveness. As use increases over time, CER becomes more feasible (green, II), and increasingly meaningful analyses can be performed in multiple subgroups (red, III).

even if balance is achieved in the small overlapping PS areas (see discussion above and **Figure 4**), there might be too few remaining comparable patients for effect estimation with any useful precision.

Disease-specific issues

In addition to confounding and precision, several other issues need to be considered that can substantially hamper phase IVA CER activities.

Long time to event. Even if many users of a newly marketed medication can be identified shortly after market access, allowing the initiation of a CE study early in phase IVA, outcomes that require long induction periods will impede the availability of timely CER results. Medications for the prevention of diseases usually fall into this category. Even large individual RCTs on the efficacy of statins used for the primary prevention of coronary heart disease showed no effect in the first 6 months after initiation of treatment. Beyond the issue of whether it is practical to study such delayed end points over a short time interval, there is another complication. The more delayed an outcome, the greater will be the extent of nonadherence to therapy, and this will need to be factored into such studies. Arguably, this may be one of the most important factors explaining differences in effect sizes between efficacy measures from highly controlled RCTs and studies that observe routine care and real-life adherence patterns.

Long exposure-effect period. The exposure-effect period is the time period during which the medication exerts an effect on a patient's physiology and produces measurable outcomes. It often starts shortly after the first tablet is taken and ends soon after the last is used. Other medications that may cause longer-term disruptions of the body's physiology may not begin to exert their effects for a substantial period of time after initiation or may have an effect that lasts far beyond the final pill. Although very long exposure-effect periods are exceptions, the exact form of the exposure-effect period generally depends on the pharmacokinetics and pharmacodynamics of the drug as well as on the outcome under study.[34] In cohort studies, the clear temporality of "exposure, then event" makes it straightforward to vary the exposure-effect period and empirically assess the most likely underlying risk period.[35]

First-in-class medications. In the context of this discussion, we use the term "first-in-class" for medications that are true innovations for an indication for which no treatment exists or only clearly inferior alternative treatments are available. In such situations, the selection pressure is even higher. Who are the patients who are still not using the innovative agent despite its clear advantages? These patients may have different indications or much less severe disease expressions. It therefore becomes difficult to find active comparator groups, and nonuser comparison groups remain suboptimal because of the increased risk of uncontrollable confounding.[36,37] In cases of rapid and almost complete market uptake of a new drug, comparisons with historical controls using time-trend analyses may be the most valid approach for estimating the added effectiveness.[37,38] Such rapid uptake is rare but may occur when a generic medication is marketed after the branded drug loses its market exclusivity and if the generic drug is more available to patients via greatly reduced copayments.[39]

Data issues

The use of secondary data, including claims data, EMRs, and certain registries, provides insights into the relationship between treatment and health outcomes in routine care. Such data are continuously collected and stored electronically as part of routine care and therefore do not perturb the care system as would randomized trials or prospective epidemiologic studies, which require patient consent and follow-up monitoring. Once patients and providers know they are being studied, or even if they are merely aware of the specific study question, they may alter their behaviors in a way that influences patient selection, treatment choice, and outcome assessment.

Despite their well-recognized advantages, secondary data come with substantial limitations.

- *Lag time:* Some secondary data, such as EMRs, are instantly recorded at the time of patient care. However, extraction of such information often takes some time. Claims data need to go through an adjudication process by insurance companies before they become useful for research. This process may take from 3 months (commercial) to a year or longer (some aspects of Medicare/Medicaid).[40]

- *Granularity of information:* In some situations, specific information with a high degree of granularity is required. In studies comparing medications, it is sometimes essential that we have details about the exact medication patients are using, including a differentiation between various brand and generic manufacturers of the same molecule, as well as the strength, quantity dispensed, dosage form, and route of administration. In the United States, this information is usually well captured with the 10-digit National Drug Code. However, because devices lack an analogous code, information on the exact type and build of a device is not captured in all recoding systems—certainly not in claims data and rarely in EMR systems.

- *Suitable outcome information:* Secondary data are well suited for capturing major medical events that lead to hospitalizations, such as stroke, myocardial infarction, surgeries, and serious infections. However, they are often limited in their ability to assess fatal events that occur outside the hospital, as well as functional status, cognitive status, pain, or quality-of-life end points that are highly relevant in studies of chronic conditions. Even if such information is routinely collected from chronically ill patients, it might not be collected at the appropriate time relative to medication exposure. For example, "baseline measurements" might already be a year old once the medication is initiated, or "follow-up measurements" might be made after the relevant exposure-risk period has passed.

- *Completeness of information:* Important confounder information is often unavailable in secondary data. Information may be missing with respect to education and socioeconomic status; lifestyle factors such as smoking and alcohol consumption, diet, and exercise patterns; body mass index; and family history of specific diseases. Even if data are meant to be recorded in EMRs, the information is not always present or the parameters may not be measured at an appropriate time. For example, investigators might want to know patients' baseline kidney function, but such data are unavailable because the tests had no relevance for the acute treatment decision and were therefore not ordered. Hence, in EMR systems we might find cholesterol levels or blood pressure measurements for some patients but not all. The presence of such information may be related to the patient's disease state; consequently, missing values cannot be considered random.[41,42]

SOME SOLUTIONS

As we noted above, payers are increasingly interested in using their own data to perform CER studies to inform decision making in the early marketing period of a new drug. Various countries are also establishing national infrastructure systems to enable near-real-time safety monitoring of medical products within the routine care setting shortly after the drugs enter the market. These systems of networked databases, such as the FDA's Sentinel System, may also serve as national resources for rapid generation of CE evidence.[43] Many of the challenges to observational CER described above can be overcome to varying extents with sound methodological approaches. However, these solutions are not fail-safe and can and should be implemented in concert with additional approaches to CER data generation and synthesis. Although the goal of CER—to understand the relative effectiveness of medical products in routine care—implies evaluation before market entry, parts of the process can be initiated prior to approval.

In this section we describe some solutions, which we present in reverse chronological order with respect to when they can be initiated, from phase IVA to phase II. We define phase IVA as the early marketing phase, in which the eventual market share and insurance-coverage status of a newly marketed drug are still in flux before a more stable postmarketing phase is reached.

Sequential cohort studies

Once a drug is on the market and enters routine practice, observational CE monitoring can begin. The balanced sequential cohort design (**Figure 7**) may become a standard solution for working with secondary observational data that fit a broad range of CE and safety questions.[19] The design is based on data collected during the process of providing care and that become available with a relatively short lag time, enabling near-real-time monitoring of effectiveness. As the drug's time on the market increases, the cohort of patients exposed to the drug will expand and can be periodically analyzed. In our opinion, sequential cohort designs are the cornerstone for a proactive approach to ascertaining the CE of newly marketed medications.

Sequential cohorts can be defined by calendar intervals as data become available, such as on a monthly or quarterly basis. In several applications of the sequential cohort approach to active drug safety monitoring, we extracted new users of either the monitored drug or an active comparator from multiple longitudinal health-care databases, in each of the calendar quarters from the time the monitored drug entered the market.[44–46] We applied relevant eligibility criteria, and then, within each quarterly period, we compared PS-matched initiators of the monitored drug with initiators of the comparator drug. We used data for the 180 days preceding each patient's index date to identify a broad range of patient characteristics for the PS models. This period necessarily extended backward into the preceding calendar quarter. We constructed separate models among new users in each calendar quarter and for each database.

Each PS model included a set of predefined covariates specific to the monitored drug and outcome. We enriched the models with empirically identified variables using the high-dimensional PS (hd-PS version 2) algorithm (available at http://hdpharmacoepi.org).[47] On the basis of on an extensive evaluation of this algorithm, we recommend variable selection based only on the covariate–exposure associations for drug comparisons wherever there is a likelihood of only a few users or outcome events in one or both of the treatment groups. With increasing numbers of users and outcome events, variable selection that takes the outcome into account will perform better.[15,48]

The result of this process is a series of PS-matched cohorts that can be conceptualized as "time-sliced" subcohorts nested within a larger open cohort study. With 1:1 matching, these quarter-by-quarter cohorts can be easily combined with cohorts from subsequent quarters as new data become available, and analyses can proceed without further adjustment for patient characteristics, calendar time, or matching sets. It is also possible to consider multiple outcomes within the same matched cohort.[50] With respect to diagnostics, the balance achieved in observable patient characteristics can easily be demonstrated by a cross-tabulation with treatment choice and with measures such as the Mahalanobis distance.[49]

Importantly, focusing on new users of the study drug and employing active comparators establish clear temporality among pretreatment variables that may confound the association between the monitored product of interest and the outcome of interest.[19,51] Balancing cohorts by pretreatment patient characteristics is a safe strategy and avoids conditioning on factors downstream from exposure.[21] The robustness and simplicity of this approach make it appealing for practical postmarketing effectiveness monitoring.

Extension of phase III and IV trials

The fastest way to collect CE data is by extending phase III trials such that participants who are already enrolled in preapproval trials can be followed into the postmarketing period. As the trial is completed and the drug enters the market, participants are sometimes asked to continue in an open-label study with free choice of medication. This effectively turns the trial into a nonrandomized registry study that collects detailed information on

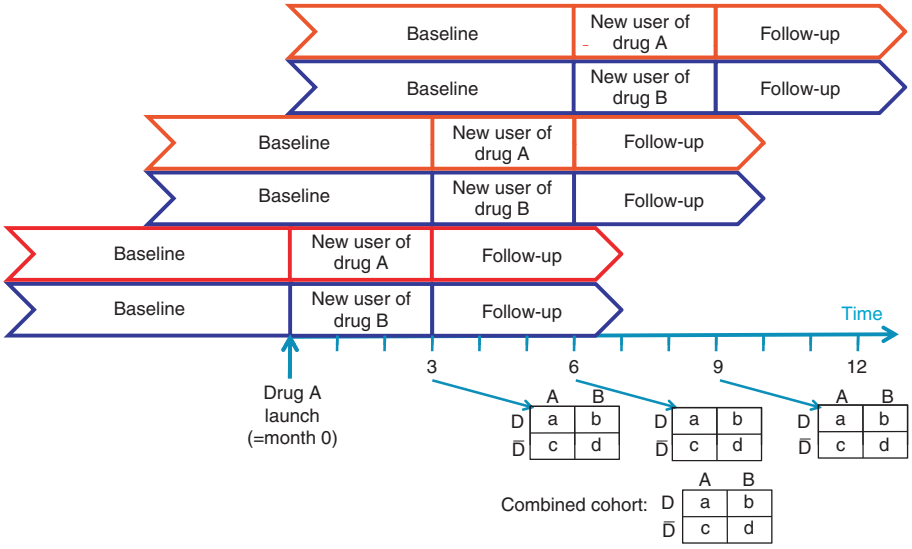

Figure 7 The balanced sequential cohort design. As additional data become available, new users of the study drug are matched to similar new users of the comparison drug, and these pairs are added to the growing cumulative cohort.

treatment choice and outcomes in a defined patient population. A major advantage of this approach is that much of the relevant information has already been collected; furthermore, the existing system of patient follow-up will accelerate the generation of CER evidence. However, generalizability will be limited depending on the phase III trial inclusion/exclusion criteria, and all limitations of nonrandomized studies noted above will also apply to this situation.

In particular, if trial participants opt to not continue in the extension study, their choice may be correlated with their risk for the outcome. The extension may allow subjects to cross over in their treatment, for example, by allowing patients in the comparator arm access to the active treatment, and factors that influence such a decision may be difficult to measure. If the trial has revealed that one treatment is clearly preferable to the other, the few subjects who remain in the other treatment group may turn out to be a selective population. In this case, their value as a comparison group would be limited. As an example, after completion of the aspirin arm of the randomized Physicians' Health Study, participants were offered the opportunity to receive active aspirin therapy. Two years later, 99% of those originally assigned to the active aspirin treatment arm were requesting active aspirin, as compared with only 75% of those who had originally been assigned to placebo. The decision to request continued active aspirin therapy was strongly related to cardiovascular risk factors in this population.[52]

When confounding is deemed insurmountable with nonrandomized studies, the only alternative is a phase IV randomized trial. Several alternative randomized designs are available to provide evidence on efficacy and safety of approved drugs. These include classic parallel-group, placebo-controlled trials to evaluate efficacy in populations not previously studied or to evaluate alternative end points. Examples include the JUPITER trial evaluating the efficacy of rosuvastatin for primary prevention in individuals with normal low-density lipoprotein cholesterol but elevated levels of high-sensitivity C-reactive protein, and trials of celecoxib for prevention of adenoma.[53,54] Although not designed with active comparators, such trials provide important additional information on efficacy and safety and can specifically identify important adverse effects such as the risk of diabetes associated with the use of statins[53] and cardiovascular complications associated with the use of celecoxib.[55]

In some instances, phase IV trials include active comparators, with the goal of evaluating either superiority or noninferiority. Examples include the VIGOR trial,[56] designed to compare the gastrotoxicity of rofecoxib with that of naproxen; PROVE-IT TIMI 22,[57] designed to evaluate the equivalence of 40 mg pravastatin vs. 80 mg atorvastatin daily in patients with acute coronary syndrome; the SEARCH trial,[58] designed to compare the efficacy of 20 mg vs. 80 mg simvastatin in individuals with a prior myocardial infarction; and the VALIANT trial,[59] designed to evaluate superiority and/or equivalence of valsartan to captopril and their combination formulation in patients with myocardial infarction complicated by left systolic dysfunction, heart failure, or both. The important information on comparative efficacy and safety provided by these trials augments what is known about these drugs from premarketing clinical trials.

Although much has been learned from these classically designed studies, concerns remain about their costs, the time they take from initial conception to dissemination of results,

and their selective populations. As a result, there is ongoing discussion about the extent to which cluster-randomized trials or encouragement trials might broaden the generalizability of randomized studies at a lower cost, while preserving important aspects of randomization.[60]

More broadly, many in the field debate the extent to which randomized trials are generalizable to routine care and whether such studies should still be considered CER. A sharp distinction between effectiveness and efficacy is usually not helpful. It is important to recognize that perfect information on drug effectiveness will rarely be available even years after market entry; we must therefore focus on determining the aspects of evidence generation that we are willing to forgo in favor of more important aspects at a given point in time and based on what we think we already know about a new drug. This is relevant not only to deciding between randomized and nonrandomized designs but also to considering the various trade-offs at any decision point in the design of a CE study. We noted above that increased validity often comes at the price of reduced generalizability, even in observational studies using secondary data.[14,18] In our opinion, the aspects that should be assigned the most weight when considering trade-offs are internal validity of findings (valid for the population studied), selection of end points important to both patients and providers, and use of clinically relevant comparison groups. Other important aspects include broader generalization, timeliness of findings, costs, and specific subgroup analyses.

Blanket statements such as "We need evidence from all types of designs available to CER" are not false.[9] However, it is more important for stakeholders to be aware that all choices between research methodologies to generate CER evidence—including choice of study design, type of analysis, data source, comparison group, and end points—come with trade-offs regarding validity, precision, timeliness, feasibility, generalizability, clinical relevance, and other attributes. As investigators, we need to aim for transparency regarding the trade-offs we make and our reasons for making them, even if we disagree about which trade-offs are more appropriate than others.

Indirect comparison of RCTs

At the time point when medications enter the market, data will already be available for the small number of efficacy trials that supported the application for regulatory approval. Most premarketing clinical trials use placebo controls, although studies relating to infection control, cancer, and select other conditions may employ active comparators. Over the past decade, of the total number of approval application packages for drugs that subsequently obtained FDA approval, 50% included a trial with an active comparator.[61] However, it remains unclear whether the comparator used and the evidence generated by those studies are relevant for routine care and therefore of value in making prescribing and coverage decisions. If no active comparator studies are available, or an unsuitable active comparator was used, then indirect comparisons may provide some further insight.

The basic aim of indirect comparisons is to identify a reference group common to a group of trials (e.g., a placebo control or a uniform active comparator) against which the efficacy of each drug of interest was assessed. Using information about the efficacy of drug A relative to a reference drug, and of drug B relative to the reference drug, we can infer the efficacy of A relative to B. However, the randomization that balanced each individual trial no longer holds because the composition of the placebo group in one trial might be quite different from that in another. This turns an indirect comparison into an epidemiological study, which requires confounding adjustment according to the jointly observed patient characteristics to stand in place of exhaustive risk factor balancing.[62] In particular, the assumption that the relative treatment effect remains constant across populations with possibly widely varying absolute risks is a very strong one, and possibly unjustifiable.

Overall, indirect comparisons have been shown to produce valid results if applied correctly.[63] Indirect comparisons can be conducted more expeditiously than *de novo* head-to-head randomized trials, and indirect comparisons using preapproval trials can be completed even before market authorization. One recent example is a comparison of dabigatran and rivaroxaban as agents to prevent venous thromboembolism in patients after knee or hip replacement.[64] Dabigatran received marketing authorization from the European Medicines Agency for the prevention of venous thromboembolism as early as March 2008 but was not approved in the United States until October 2010. The indirect comparison was completed on 28 September 2009, before any head-to-head trial was completed. Enoxaparin was used as the common reference group because it had been compared with dabigatran in three trials and with rivaroxaban in six. Rivaroxaban was found to be more effective than either enoxaparin or dabigatran.

Indirect comparisons can be extended to networks of randomized trials as more evidence becomes available,[65,66] although they are often hindered by the lack of common reference populations among preapproval trials or by the lack of comparability in patient inclusion and end-point definitions of various trials. Furthermore, although indirect comparisons are, in practice, based almost exclusively on randomized efficacy trials, in theory this is not necessary. This is more a consequence of the available data than of the intention. As with all of our proposed solutions, trade-offs must be made between generalizability to routine care and the time until evidence becomes available.

Modeling and trial simulation

Pharmacokinetic and pharmacodynamic modeling has long guided clinical trial design, dose selection, and development strategy to increase chances of identifying efficacious therapeutic options and moving them into experimental testing.[67,68] Modeling techniques have become increasingly sophisticated,[69] and their use to enhance sponsors' ability to predict desired outcomes in prospective trials may help accelerate drug development and prioritize product development. However, such modeling does not fully simulate a virtual randomized trial to assess the existence and magnitude of an effect.

Computer modeling of human physiology and the effects of biologically active molecules can be more powerful than pharmacokinetic and pharmacodynamic approaches. Simulation

software tools such as Archimedes[70] mathematically represent physiological pathways and the effects of multiple diseases, tests, and treatments, using hundreds of differential equations. These models even include parameters on health-care practice and administrative events so that the effectiveness of intervention in routine care can be estimated. The flexibility of the system allows testing of a wide variety of interventions, including multiple simultaneous interventions, as well as a large set of clinically meaningful outcomes, including clinical outcomes such as myocardial infarctions and quality-adjusted life years. In numerous validation studies comparing the Archimedes simulations against the findings of randomized trials, the developers of Archimedes claim a correlation in effect estimates of 0.96.[71] More independent evaluations may be necessary, although the existing evidence is striking. Such powerful simulation studies may be the very first evidence generated on the CE of drugs, even before clinical trials are completed.[72] Some health plans, such as Kaiser Permanente, already use Archimedes to inform treatment-guideline and coverage decisions in the early marketing phase.[73]

Other modeling approaches are based on extrapolating findings from selected studies to populations that were not included or to long-term effects that were not observable. Such models may also be useful when the effectiveness of multiple clinical options needs to be compared and when it is difficult to conduct a single large study.[74]

IMPLICATIONS FOR DRUG DEVELOPMENT

Mastering the methods described in the previous sections will become one of the most critical preconditions for the success of pharmaceutical companies in preparing their development and market access capability against the backdrop of the shifting regulatory and reimbursement paradigm. Of particular relevance for the industry, this shift is driven by three major changes and their interplay:

1. The increasing influence of payers on prescribing decisions by the individual physician and on driving new, real-world evidence standards to which reimbursement/pricing and utilization will be tied. This trend carries the risk of slowing the process of bringing new drugs to market. This risk and its consequences can be best mitigated through the systematic development of robust CE data, given that initial indications

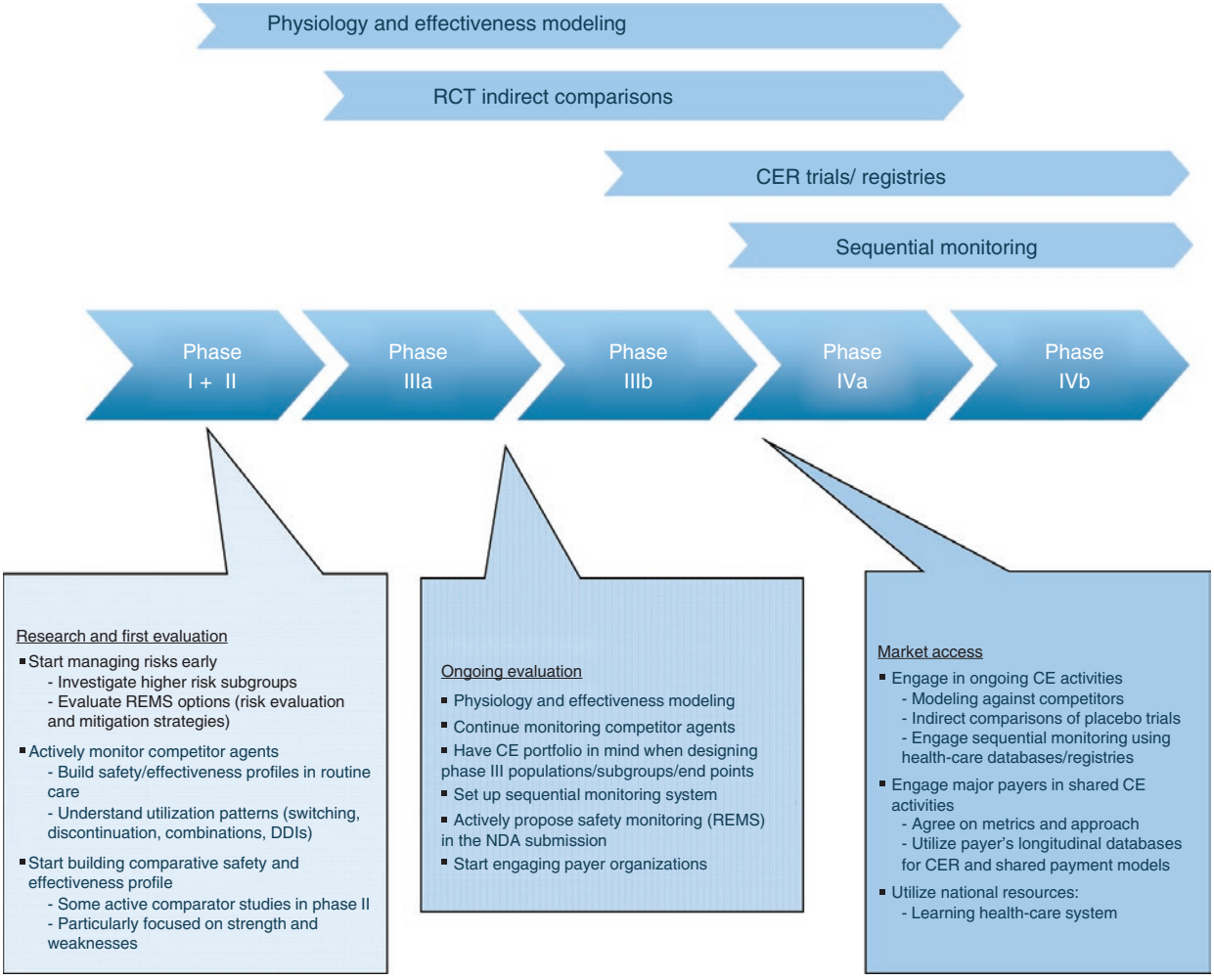

Figure 8 Comparative-effectiveness research (CER) in drug development and market access. DDI, drug–drug interaction; NDA, new drug application; RCT, randomized controlled trial.

and coverage decisions are regularly reviewed and modified based on accumulating postmarketing evidence about the added benefits and/or reduced risks of new medications.
2. Increasingly strong coordination between payers and regulatory agencies, particularly regarding evidence standards. For example, the FDA's Mini-Sentinel system involves many of the most relevant U.S. payer organizations as data partners. This development reinforces the impact of payers and, through regulatory mechanisms such as conditional approval decisions and restriction of the initial label to small patient populations, puts pharmaceutical companies at risk of facing slowdowns in future market penetration or strategic expansion of new medications into new segments of the population. High penetration and expansion can still be achieved with early and robust CE data paired with proactive risk-minimization strategies.
3. An accelerating trend toward CE evidence generation by third parties, driven largely by insurers and provider organizations leveraging the wealth of data from their longitudinal claims or EMR systems, as well as academic groups benefiting from increased governmental funding for CER. This trend will reduce pharmaceutical companies' control over the flow of evidence regarding their own products and challenge them to develop and validate postmarketing evidence faster than any third party.

In an effort to mitigate the risks posed by these changes and transform them into a source of sustainable competitive advantage, we present the solutions discussed above as specific recommendations in relation to the typical medication life cycle (**Figure 8**). We also recommend that pharmaceutical companies embark on two mutually reinforcing strategic moves:

1. First, to prepare their organizations to drive the transition from the traditional regulatory "event" model—in which a "one-time" definitive and appropriately broad approval and reimbursement decision was made by regulators, payers, and other reimbursement stakeholders based on placebo-controlled clinical trial data—toward a "process" model, in which reimbursement stakeholders condition their initial approval and reimbursement decision on the continued development of CE and other scientific evidence in situations in which significant uncertainty around the health outcomes of medications remains at the time of marketing authorization.
2. Second, to build a system of capabilities supporting the seamless access, integration, and analytics of routine-care data before and after market access and to align such a system with existing procedures for generating and assessing clinical trial data. This may require new types of integrated research alliances between pharmaceutical manufacturers and payers.

A major step toward building such a supporting system of capabilities and the key to strategic use of real-world data is to transform CER from an unreliable, *ad hoc* approach to a robust science that minimizes bias and is scalable across the needs of clinicians, regulators, payers, and the industry. A strategy exploiting the various solutions described in this article, either simultaneously or sequentially (**Figure 8**), may open the door to truly strategic use of CER.

Pharmaceutical companies could begin with the specific step of determining what an effective system of capabilities for generating and using relevant real-world data looks like, followed by conscious investment in building it into a major driver of competitive advantage. Next, investment in demonstration projects using the methods described in this article would be a pragmatic approach to achieving early access to all relevant data and capabilities.

CONCLUSION

Establishing the CE of newly marketed medications is challenging but crucial: during the critical early postmarketing period, decisions are made regarding a drug's added therapeutic value for patients. These decisions, although often made with limited CER evidence, serve as the basis for new utilization patterns that are quickly established but difficult to change. Putting the wrong pattern in place sets in motion the attendant consequences of health and safety for patients and of cost for health-care providers. Robust CE results, produced as quickly as possible, have enormous importance for patients and payers.

No single approach will fulfill all possible needs for CE information shortly after a drug has entered the market. We propose a mix of approaches that includes sequential cohort monitoring with secondary health-care data, phase III RCT spillover studies and phase IV trials, indirect comparisons of placebo-controlled trials, and modeling and simulating virtual trials.

The issues we discuss here are highly relevant for drug development. It is increasingly important to effectively integrate information from a variety of sources, multiple partners, and different development stages to produce a successful CER portfolio in the early marketing period. However, this may call for creation of integrated CER analytics capabilities within companies.

With the emergence and integration of CER data throughout the development process, it is vital for all stakeholders to understand that the various research methodologies to generate CER evidence—including choice of study design, type of analysis, data source, comparison group, and end points—have trade-offs with respect to validity, precision, timeliness, feasibility, generalizability, clinical relevance, and other aspects. We need to aim for transparency regarding the trade-offs we make and why we make them, even if we disagree about their relevance and value.

ACKNOWLEDGMENTS

S.S. is a member of the Methods Committee of the Patient-Centered Outcomes Research Institute and director of the Brigham and Women's Hospital DEcIDE Center for CER and the DEcIDE Methods Center, both funded by the Agency for Healthcare Research and Quality (AHRQ). The opinions expressed here are only those of the authors and not necessarily those of the aforementioned institutions. S.S.'s work was funded by an investigator-initiated research grant from Novartis Pharmaceuticals Corporation and was also supported by grants from the National Library of Medicine (RO1-LM010213 and RC1-LM010351), the National Heart, Lung,

and Blood Institute (RC4-HL102023), and the National Center for Research Resources (RC1-RR028231). J.A.R. was supported by a career development award from the AHRQ (K01-HS018088). The authors acknowledge productive discussions with Jens Grueger, Richard Willke (both of Pfizer Inc.), Kristijan Kahler (Novartis Pharmaceuticals Corporation), and Adrian Towse (Office of Health Economics) that helped shape this article.

CONFLICT OF INTEREST

The authors declared no conflict of interest.

1. Schneeweiss, S. Developments in post-marketing comparative effectiveness research. *Clin. Pharmacol. Ther.* **82**, 143–156 (2007).
2. Goldberg, N.H., Schneeweiss, S., Kowal, M.K. & Gagne, J.J. Availability of comparative efficacy data at the time of drug approval in the United States. *JAMA* **305**, 1786–1789 (2011).
3. Califf, R.M. *et al*. An update on the IMProved reduction of outcomes: Vytorin Efficacy International Trial (IMPROVE-IT) design. *Am. Heart J.* **159**, 705–709 (2010).
4. Jackevicius, C.A., Tu, J.V., Ross, J.S., Ko, D.T. & Krumholz, H.M. Use of ezetimibe in the United States and Canada. *N. Engl. J. Med.* **358**, 1819–1828 (2008).
5. Sorenson, C. Use of comparative effectiveness research in drug coverage and pricing decisions: a six-country comparison. *Issue Brief (Commonw. Fund)* **91**, 1–14 (2010).
6. US Food and Drug Administration. MOU 225-10-0010: Memorandum of understanding between US Food and Drug Administration and Centers for Medicare and Medicaid Services <http://www.fda.gov/aboutfda/partnershipscollaborations/memorandaofunderstandingmous/domesticmous/ucm217585.htm> (2010).
7. Cioffe, C. Portfolio selection and management in pharmaceutical research and development: issues and challenges. *Clin. Pharmacol. Ther.* **89**, 300–303 (2011).
8. Honig, P.K. Comparative effectiveness: the fourth hurdle in drug development and a role for clinical pharmacology. *Clin. Pharmacol. Ther.* **89**, 151–156 (2011).
9. Velentgas, P.*et al*. A survey of aspirin use, non-prescription NSAID use, and cigarette smoking among users and non-users of prescription NSAIDs:estimates of the effect of unmeasured counfounding by these factors on studies of NSAID use and risk of myocardial infarction. *Pharmacoepidemiol. Drug Saf.* **10**, S103 (2001).
10. Schneeweiss, S. & Avorn, J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J. Clin. Epidemiol.* **58**, 323–337 (2005).
11. Schneeweiss, S., Glynn, R.J., Avorn, J. & Solomon, D.H. A Medicare database review found that physician preferences increasingly outweighed patient characteristics as determinants of first-time prescriptions for COX-2 inhibitors. *J. Clin. Epidemiol.* **58**, 98–102 (2005).
12. Tan, H. *et al*. Impact of asthma controller medications on clinical, economic, and patient-reported outcomes. *Mayo Clin. Proc.* **84**, 675–684 (2009).
13. Segal, J.B., Dy, S.M., Millman, E.A., Herbert, R., Bass, E.B. & Wu, A. Diffusion into use of exenatide for glucose control in diabetes mellitus: a retrospective cohort study of a new therapy. *Clin. Ther.* **29**, 1784–1794 (2007).
14. Schneeweiss, S. *et al*. Increasing levels of restriction in pharmacoepidemiologic database studies of elderly and comparison with randomized trial results. *Med. Care* **45**, S131–S142 (2007).
15. Glynn, R.J., Schneeweiss, S. & Sturmer, T. Propensity scores in pharmacoepidemiology. *Basic Clin. Phar. Tox.* **98**, 252–259 (2006).
16. Cochrane, A. *Effectiveness and Efficiency: Random Reflection on Health Services* (Nuffield Provincial Trust, London, 1972).
17. Concato, J., Shah, N. & Horwitz, R.I. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N. Engl. J. Med.* **342**, 1887–1892 (2000).
18. Psaty, B.M. & Siscovick, D.S. Minimizing bias due to confounding by indication in comparative effectiveness research: the importance of restriction. *JAMA* **304**, 897–898 (2010).
19. Schneeweiss, S. A basic study design for expedited safety signal evaluation based on electronic healthcare data. *Pharmacoepidemiol. Drug Saf.* **19**, 858–868 (2010).
20. Rubin, D.B. & Thomas, N. Combining propensity score matching with additional adjustments for prognostic covariates. *J. Am. Stat. Assoc.* **95**, 573–585 (2000).
21. Myers, J. *et al*. Effects of adjusting for instrumental variables on bias and precision of effect estimates. *Am. J. Epidemiol.* (in press).
22. Brookhart, M.A., Wang, P.S., Solomon, D.H. & Schneeweiss, S. Evaluating short-term drug effects using a physician-specific prescribing preference as an instrumental variable. *Epidemiology* **17**, 268–275 (2006).
23. Brookhart, M.A. & Schneeweiss, S. Preference-based instrumental variable methods for the estimation of treatment effects: assessing validity and interpreting results. *Int. J. Biostat.* **3**, 14 (2007).
24. Rassen, J.A., Schneeweiss, S., Glynn, R.J., Mittleman, M.A. & Brookhart, M.A. Instrumental variable analysis for estimation of treatment effects with dichotomous outcomes. *Am. J. Epidemiol.* **169**, 273–284 (2009).
25. Tamblyn, R., McLeod, P., Hanley, J.A., Girard, N. & Hurley, J. Physician and practice characteristics associated with the early utilization of new prescription drugs. *Med. Care* **41**, 895–908 (2003).
26. Brookhart, M.A., Rassen, J.A., Wang, P.S., Dormuth, C., Mogun, H. & Schneeweiss, S. Evaluating the validity of an instrumental variable study of neuroleptics: can between-physician differences in prescribing patterns be used to estimate treatment effects? *Med. Care* **45**, S116–S122 (2007).
27. Mintzes, B. *et al*. Influence of direct to consumer pharmaceutical advertising and patients' requests on prescribing decisions: two site cross sectional survey. *BMJ* **324**, 278–279 (2002).
28. Mintzes, B. *et al*. How does direct-to-consumer advertising (DTCA) affect prescribing? A survey in primary care environments with and without legal DTCA. *CMAJ* **169**, 405–412 (2003).
29. Schneeweiss, S., Patrick, A.R., Maclure, M., Dormuth, C.R. & Glynn, R.J. Adherence to statin therapy under drug cost sharing in patients with and without acute myocardial infarction: a population-based natural experiment. *Circulation* **115**, 2128–2135 (2007).
30. Peduzzi, P., Concato, J., Kemper, E., Holford, T.R. & Feinstein, A.R. A simulation study of the number of events per variable in logistic regression analysis. *J. Clin. Epidemiol.* **49**, 1373–1379 (1996).
31. Cepeda, M.S., Boston, R., Farrar, J.T. & Strom, B.L. Comparison of logistic regression versus propensity score when the number of events is low and there are multiple confounders. *Am. J. Epidemiol.* **158**, 280–287 (2003).
32. Rubin, D.B. Estimating causal effects from large data sets using propensity scores. *Ann. Intern. Med.* **127**, 757–763 (1997).
33. Rassen, J.A., Glynn, R.J., Rothman, K.J., Setoguchi, S. & Schneeweiss, S. Applying propensity scores estimated in a full cohort to adjust for confounding in subgroup analyses. *Pharmacoepidemiol. Drug Saf.* (in press).
34. Solomon, D.H., Lunt, M. & Schneeweiss, S. The risk of infection associated with tumor necrosis factor alpha antagonists: making sense of epidemiologic evidence. *Arthritis Rheum.* **58**, 919–928 (2008).
35. McMahon, A.D., Evans, J.M., McGilchrist, M.M., McDevitt, D.G. & MacDonald, T.M. Drug exposure risk windows and unexposed comparator groups for cohort studies in pharmacoepidemiology. *Pharmacoepidemiol. Drug Saf.* **7**, 275–280 (1998).
36. Rosenbaum, P.R. Differential effects and generic biases in observational studies. *Biometrika* **93**, 573–586 (2006).
37. Brookhart, M.A., Rassen, J.A. & Schneeweiss, S. Instrumental variable methods in comparative safety and effectiveness research. *Pharmacoepidemiol. Drug Saf.* **19**, 537–554 (2010).
38. Schneeweiss, S., Maclure, M., Soumerai, S.B., Walker, A.M. & Glynn, R.J. Quasi-experimental longitudinal designs to evaluate drug benefit policy changes with low policy compliance. *J. Clin. Epidemiol.* **55**, 833–841 (2002).
39. Paterson, J.M., Mamdani, M., Juurlink, D.N., Naglie, G., Laupacis, A. & Stukel, T.A. Clinical consequences of generic warfarin substitution: an ecological study. *JAMA* **296**, 1969–1972 (2006).
40. Hartzema, A.G., Racoosin, J.A., MaCurdy, T.E., Gibbs, J.M. & Kelman, J.A. Utilizing Medicare claims data for real-time drug safety evaluations:is it feasible? *Pharmacoepidemiol. Drug Saf.* **20**, 684–688 (2011).
41. Hamilton, W., Lancashire, R., Sharp, D., Peters, T.J., Cheng, K.K. & Marshall, T. The importance of anaemia in diagnosing colorectal cancer: a case-control study using electronic primary care records. *Br. J. Cancer* **98**, 323–327 (2008).
42. Mulla, Z.D., Seo, B., Kalamegham, R. & Nuwayhid, B.S. Multiple imputation for missing laboratory data: an example from infectious disease epidemiology. *Ann. Epidemiol.* **19**, 908–914 (2009).
43. Behrman, R.E., Benner, J.S., Brown, J.S., McClellan, M., Woodcock, J. & Platt, R. Developing the Sentinel System–a national resource for evidence development. *N. Engl. J. Med.* **364**, 498–499 (2011).
44. Seeger, J.D., Walker, A.M., Williams, P.L., Saperia, G.M. & Sacks, F.M. A propensity score-matched cohort study of the effect of statins, mainly fluvastatin, on the occurrence of acute myocardial infarction. *Am. J. Cardiol.* **92**, 1447–1451 (2003).

45. Seeger, J.D., Kurth, T. & Walker, A.M. Use of propensity score technique to account for exposure-related covariates: an example and lesson. *Med. Care* **45**, S143–S148 (2007).
46. Gagne, J.J., Glynn, R.J., Walker, A.M., Glynn, R.J. & Schneeweiss, S. Comparison of alerting algorithms for prospective medical product safety monitoring: a simulation study and empirical application. *Pharmacoepidemiol. Drug Saf.* **20**, S143–S144 (2011).
47. Schneeweiss, S., Rassen, J.A., Glynn, R.J., Avorn, J., Mogun, H. & Brookhart, M.A. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. *Epidemiology* **20**, 512–522 (2009).
48. Rassen, J.A., Glynn, R.J., Brookhart, M.A. & Schneeweiss, S. Covariate selection in high-dimensional propensity score analyses of treatment effects in small samples. *Am. J. Epidemiol.* **173**, 1404–1413 (2011).
49. Mahalanobis, P.C. On the generalized distance in statistics. *Proc. Natl. Inst. Sci. India* **12**, 49–55 (1936).
50. Rassen, J.A., Solomon, D.H., Glynn, R.J. & Schneeweiss, S. Simultaneously assessing intended and unintended treatment effects of multiple treatment options: a pragmatic "matrix design." *Pharmacoepidemiol. Drug Saf.* **20**, 675–683 (2011).
51. Ray, W.A. Evaluating medication effects outside of clinical trials: new-user designs. *Am. J. Epidemiol.* **158**, 915–920 (2003).
52. Cook, N.R., Hebert, P.R., Manson, J.E., Buring, J.E. & Hennekens, C.H. Self-selected posttrial aspirin use and subsequent cardiovascular disease and mortality in the physicians' health study. *Arch. Intern. Med.* **160**, 921–928 (2000).
53. Ridker, P.M. *et al.*; JUPITER Study Group. Rosuvastatin to prevent vascular events in men and women with elevated C-reactive protein. *N. Engl. J. Med.* **359**, 2195–2207 (2008).
54. Arber, N. *et al.*; PreSAP Trial Investigators. Celecoxib for the prevention of colorectal adenomatous polyps. *N. Engl. J. Med.* **355**, 885–895 (2006).
55. Solomon, S.D. *et al.*; APC and PreSAP Trial Investigators. Effect of celecoxib on cardiovascular events and blood pressure in two trials for the prevention of colorectal adenomas. *Circulation* **114**, 1028–1035 (2006).
56. Bombardier, C. *et al.*; VIGOR Study Group. Comparison of upper gastrointestinal toxicity of rofecoxib and naproxen in patients with rheumatoid arthritis. VIGOR Study Group. *N. Engl. J. Med.* **343**, 1520–1528, 2 p following 1528 (2000).
57. Cannon, C.P. *et al.*; Pravastatin or Atorvastatin Evaluation and Infection Therapy-Thrombolysis in Myocardial Infarction 22 Investigators. Intensive versus moderate lipid lowering with statins after acute coronary syndromes. *N. Engl. J. Med.* **350**, 1495–1504 (2004).
58. Armitage, J. *et al.*; Study of the Effectiveness of Additional Reductions in Cholesterol and Homocysteine (SEARCH) Collaborative Group. Intensive lowering of LDL cholesterol with 80 mg versus 20 mg simvastatin daily in 12,064 survivors of myocardial infarction: a double-blind randomised trial. *Lancet* **376**, 1658–1669 (2010).
59. White, H.D. *et al.*; VALIANT Investigators. Mortality and morbidity remain high despite captopril and/or valsartan therapy in elderly patients with left ventricular systolic dysfunction, heart failure, or both after acute myocardial infarction: results from the Valsartan in Acute Myocardial Infarction Trial (VALIANT). *Circulation* **112**, 3391–3399 (2005).
60. Tunis, S.R., Stryer, D.B. & Clancy, C.M. Practical clinical trials: increasing the value of clinical research for decision making in clinical and health policy. *JAMA* **290**, 1624–1632 (2003).
61. McCormick, D., Gurwitz, J.H., Lessard, D., Yarzebski, J., Gore, J.M. & Goldberg, R.J. Use of aspirin, beta-blockers, and lipid-lowering medications before recurrent acute myocardial infarction: missed opportunities for prevention? *Arch. Intern. Med.* **159**, 561–567 (1999).
62. Song, F., Loke, Y.K., Walsh, T., Glenny, A.M., Eastwood, A.J. & Altman, D.G. Methodological problems in the use of indirect comparisons for evaluating healthcare interventions: survey of published systematic reviews. *BMJ* **338**, b1147 (2009).
63. Bucher, H.C., Guyatt, G.H., Griffith, L.E. & Walter, S.D. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *J. Clin. Epidemiol.* **50**, 683–691 (1997).
64. Loke, Y.K. & Kwok, C.S. Dabigatran and rivaroxaban for prevention of venous thromboembolism–systematic review and adjusted indirect comparison. *J. Clin. Pharm. Ther.* **36**, 111–124 (2011).
65. Singh, J.A. *et al.* A network meta-analysis of randomized controlled trials of biologics for rheumatoid arthritis: a Cochrane overview. *CMAJ* **181**, 787–796 (2009).
66. Salliot, C. *et al.* Indirect comparisons of the efficacy of biological antirheumatic agents in rheumatoid arthritis in patients with an inadequate response to conventional disease-modifying antirheumatic drugs or to an anti-tumour necrosis factor agent: a meta-analysis. *Ann. Rheum. Dis.* **70**, 266–271 (2011).
67. Breimer, D.D. & Danhof, M. Relevance of the application of pharmacokinetic-pharmacodynamic modelling concepts in drug development. The "wooden shoe' paradigm. *Clin. Pharmacokinet.* **32**, 259–267 (1997).
68. Danhof, M., Alvan, G., Dahl, S.G., Kuhlmann, J. & Paintaud, G. Mechanism-based pharmacokinetic-pharmacodynamic modeling-a new classification of biomarkers. *Pharm. Res.* **22**, 1432–1437 (2005).
69. Holford, N., Ma, S.C. & Ploeger, B.A. Clinical trial simulation: a review. *Clin. Pharmacol. Ther.* **88**, 166–182 (2010).
70. Archimedes. Archimedes: Quantifying Healthcare <http://archimedesmodel.com>.
71. Eddy, D.M. & Schlessinger, L. Validation of the Archimedes diabetes model. *Diabetes Care* **26**, 3102–3110 (2003).
72. Peskin, B.R., Shcheprov, A.V., Boye, K.S., Bruce, S., Maggs, D.G. & Gaebler, J.A. Cardiovascular outcomes associated with a new once-weekly GLP-1 receptor agonist vs. traditional therapies for type 2 diabetes: a simulation analysis. *Diabetes. Obes. Metab.* **13**, 921–927 (2011).
73. Dudl, R.J., Wang, M.C., Wong, M. & Bellows, J. Preventing myocardial infarction and stroke with a simplified bundle of cardioprotective medications. *Am. J. Manag. Care* **15**, e88–e94 (2009).
74. Hlatky, M.A. What we can learn from a decision model: comment on "Cost-effectiveness of adding magnetic resonance imaging to rheumatoid arthritis management." *Arch. Intern. Med.* **171**, 667–668 (2011).