Real World Data in Adaptive Biomedical Innovation: A Framework for Generating Evidence Fit for Decision-Making

S Schneeweiss[1], H-G Eichler[2], A Garcia-Altes[3], C Chinn[4], A-V Eggimann[5], S Garner[6], W Goettsch[7], R Lim[8], W Löbker[9], D Martin[10], T Müller[11], BJ Park[12], R Platt[13], S Priddy[14], M Ruhl[15], A Spooner[16], B Vannieuwenhuyse[17] and RJ Willke[18]

Analyses of healthcare databases (claims, electronic health records [EHRs]) are useful supplements to clinical trials for generating evidence on the effectiveness, harm, use, and value of medical products in routine care. A constant stream of data from the routine operation of modern healthcare systems, which can be analyzed in rapid cycles, enables incremental evidence development to support accelerated and appropriate access to innovative medicines. Evidentiary needs by regulators, Health Technology Assessment, payers, clinicians, and patients after marketing authorization comprise (1) monitoring of medication performance in routine care, including the materialized effectiveness, harm, and value; (2) identifying new patient strata with added value or unacceptable harms; and (3) monitoring targeted utilization. Adaptive biomedical innovation (ABI) with rapid cycle database analytics is successfully enabled if evidence is meaningful, valid, expedited, and transparent. These principles will bring rigor and credibility to current efforts to increase research efficiency while upholding evidentiary standards required for effective decision-making in healthcare.

Patients, particularly those with conditions that currently have limited treatment options, desire faster access to new medications that are safe and promise to reduce their suffering. Adaptive biomedical innovation (ABI) is a "multistakeholder-driven approach to improving the effectiveness and sustainability of biomedical innovation by accelerating clinical value delivery to patients while continuously improving the knowledge of new medical treatments" (Hirsch G. *et al*. in this issue).

Some new medications are produced based on an unprecedented understanding of the underlying biology and target a highly selected patient group by severity or genomic markers, providing regulators with greater confidence to approve medications with less clinical experience earlier than traditional approval pathways. Independent of the specific regulatory pathway, including accelerated approval, adaptive pathways, etc.,[1] any effort to bring innovative medications to patients faster requires (1) decision-making with less certainty at the time of initial licensing and (2) an explicit plan to quickly reduce uncertainty after initial licensing for a targeted patient population and later possibly expanding use to follow-up populations by closely monitoring the effectiveness, harm, utilization, and value in rapid analysis cycles.

The requirements and strategies for evidence development throughout this process are quite variable and depend on the condition, the target population, the drug, and available treatment alternatives. The process is further complicated by the fact that regulators and payers may have different evidence needs for

[1]Division of Pharmacoepidemiology (DoPE), Department of Medicine, Brigham & Women's Hospital, Boston, Massachusetts, USA; [2]European Medicines Agency (EMA), London, United Kingdom; [3]Agència de Qualitat i Avaluació Sanitàries de Catalunya (AQuAS), Barcelona, Spain; [4]Sanofi, Paris, France; [5]Bluebird Bio, Inc., Cambridge, Massachusetts, USA; [6]National Institute for Health and Care Excellence (NICE), London, United Kingdom; [7]National Health Care Institute, Diemen and Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, Utrecht, The Netherlands; [8]Health Products and Food Branch, Health Canada, Ottawa, Ontario, Canada; [9]Gemeinsamer Bundesausschuss (GBA), Abteilung Arzneimittel, Berlin, Germany; [10]Center for Drug Evaluation and Research, U.S. Food and Drug Administration, Silver Spring, Maryland, USA; [11]Gemeinsamer Bundesausschuss (GBA), Abteilung Arzneimittel, Berlin, Germany; [12]Seoul National University, College of Medicine, Department of Preventive Medicine, Seoul, South Korea; [13]Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Healthcare Institute, Boston, Massachusetts, USA; [14]Comprehensive Health Insights (CHI), Humana, Louisville, Kentucky, USA; [15]Aetion Inc., New York, NY, USA; [16]Health Products Regulatory Authority (HPRA), Dublin, Ireland; [17]Innovative Medicine Initiative – European Medical Information Framework, Janssen Pharmaceutica Research and Development, Beerse, Belgium; [18]International Society for Pharmacoeconomics and Outcomes Research, Lawrenceville, New Jersey, USA. Correspondence: S Schneeweiss (schneeweiss@post.harvard.edu)

Received 6 July 2016; accepted 13 September 2016; advance online publication 14 September 2016. doi:10.1002/cpt.512

Box 1

Health care databases for assessing use, effectiveness, and safety of drugs

Health care databases are produced by the operation of health care systems. They may contain information provided by physicians and other health care providers documenting the provision of care, which is typically stored in EHRs. Other databases are recordings of financial transactions and the administration of care, including claims data, pharmacy dispensing, hospital and physician billing, medical devices implantation, etc. Health care databases are considered part of RWD.

A key advantage of health care databases is that they record the routine operation of a system. Patients and providers are not aware that they might be studied using these data. Utilization and financial data reflect the reality of routine care. Therefore, the insights that can be gained from such data are directly applicable to a system that provides clinical care under usual circumstances. This might not always be optimal care and terms of who receives what treatment (e.g., off-label use) and how treatment is delivered (e.g., nonadherence) compared to a highly controlled trial.

A key limitation of health care databases is related to its advantage, the fact that data were recorded for operating a health care system and not for research. In addition to patient selection into treatment groups, some relevant information may be miscoded, incompletely recorded, or not recorded at all. This may lead to biased treatment effect estimates through residual confounding and exposure/outcome misclassification.[70] Although methodologies to overcome these limitations have improved substantially, there may be situations in which bias may represent a challenge in interpreting findings.

deciding on licensing and coverage and different timelines for formulary positioning. The ideal evidence-development plan would combine a number of study designs with or without baseline randomization and various data types (primary data collection, secondary health care databases, registries, etc.) in a complementary way to maximize the understanding of a drug's net benefit in an expedited and foreseeable yet adaptive way. Key evidentiary needs during the postmarketing phase include: (1) monitoring of medication performance in routine care, including materialized effectiveness, harm, and value; (2) identifying new patient strata in which a drug may have added value or unacceptable harm; and (3) monitoring targeted utilization and evaluation of utilization control mechanisms.[2]

Healthcare databases are longitudinal patient-level data produced by the operation of modern healthcare systems (see **Box 1**). Such real world data (RWD) are frequently used in assessing the added value of medications after marketing. A key advantage of health care databases is that they record utilization, health events, and financial data in routine care. A key limitation of health care databases is that data were not recorded for research. In addition to patient selection into treatment groups, some relevant information may be miscoded, incompletely recorded, or not recorded at all. To be useful for research the data need to be fit for the specified purpose.

Focusing on health care databases and their analysis in rapid cycles, we identify four key requirements to satisfy the evidence needs for successful accelerated-approval pathways. We seek to outline the choices and tradeoffs when considering health care databases for postmarketing monitoring activities to support ABI. This article illustrates the decision-making process with selected examples and addresses external constraints imposed by local, regional, or national regulations. We conclude by identifying areas in which health care databases and rapid-cycle analytics currently have the greatest value, useful regulatory and coverage adjustments in the near future, and recommended longer-term legislative changes.

Evidence requirements from routine health care data to support ABIs

The analysis of electronic data streams collected during the provision of healthcare has been developing over the past 25 years[3] and has begun to reflect recurrent principles recognized by researchers,[4] regulators,[5,6] and Health Technology Assessment/payers.[7] However, despite advances, there is still tremendous variation in the transparent conduct and validity of RWD analyses.[8]

To meaningfully support incremental decision-making for ABI, several principles make evidence from RWD analyses more likely to be useful for understanding the materialized effectiveness, harm, and value, identifying new patient strata with added value and monitoring targeted utilization. Four key principles of successful RWD analyses make up "MVET"[9]:

1. Meaningful evidence: Relevant and context-informed evidence based on fit-for-purpose data sufficient for interpretation and making decisions.
2. Valid evidence: Evidence that meets scientific and technical quality standards to allow causal interpretations.
3. Expedited evidence: Incremental evidence generation that is synchronized with timely decision-making.
4. Transparent evidence: Evidence that is produced transparently and therefore becomes reproducible, replicable, and trusted by decision-makers.

Evidence that meets the MVET requirements is expected to be above a critical threshold for meaningful decision-making, as laid out previously.[9] At this point, MVET can be seen as a tool to structure and clarify discussions around RWD analysis. It will help the stakeholder and research communities to converge on the best possible solution to study the net health benefit of an innovative product in a specific context. **Supplementary Table S1** online provides a worksheet to assess the MVET criteria in specific settings. In contrast to existing guidance, documents that provide more depth on processes to implement

Table 1 Frequently suggested study/data types based on routine data collections in support of ABI

	Relevance to adaptive innovation	Limitation
Study options based on routine data for adaptive innovation		
• Claims data studies	• Reflects routine care with major clinical endpoints, utilization, costs • Most complete capture of professional care encounters	Lack of clinical detail, endpoint adjudication not always feasible, no systematic assessment of clinical conditions, no PROs
• EHR study	• More clinical details • Opportunity to reach back to patients for additional information	Frequently missing data, no systematic assessment of conditions, limited PROs
• Disease registry studies	• Systematical assessment of investigator-designed measurements of relevant clinical parameters • Capture of PROs	Limited availability, time-consuming to initiate, may not represent routine care, no relevant cost outcomes
Likely less valuable options for adaptive innovation		
• Exposure-based registries	Easier to implement and often aligned with current regulator/payer requirements	Complete lack of comparators may reduce the utility of evidence
• *De novo* cohort studies	Can be highly targeted to address specific knowledge gaps	In most cases, too time-consuming to be meaningful for adaptive innovation

The focus of this enumeration is on the scientific principles, not the exact regulatory definition, although they are largely aligned. ABI, adaptive biomedical innovation; HER, electronic health records; PRO, patient reported outcomes.

database studies,[4,5,7,10] MVET is intended to provide targeted guidance for decision-makers on how to effectively use health care database analytics for incremental evidence generation supporting ABIs.

Meaningful evidence

In order for evidence for health care databases to be meaningful, it needs to contain the relevant information of sufficient quality to answer a specific question even if this information is found selectively in disparate data sources, data access should be such that ongoing monitoring is possible, and analyses need to provide measures relevant for decision-makers.

Finding the most suitable data sources for a study question

For any health care database analysis to be meaningful, of course, the appropriate information to answer the central question at hand needs to be available. It does neither mean that information needs to be comprehensive all the time nor that it needs to be perfect in all aspects. However, in most cases, this should include reliable information on drug exposure, outcomes that matter to patients, providers, regulators, and payers, and measurement of important confounders or proxies thereof. The accurate recording of dates is critical to establish temporality. Increasingly, biomarker information and other disease subgroup indicators to identify patients most likely to benefit and/or experience harm from highly targeted therapeutics (**Table 1**).

Unfortunately, there is no generalizable advice regarding where optimal information for a given question can be found. Today, most health care systems have some mechanism of electronically capturing all professional services provided, including diagnostic and procedural information. This is often linked to payments and aggregated into claims databases. Such information is more reliable in areas in which audits are frequent (hospitals) or fraudulent charging practices may lead to lawsuits (procedures). Pharmacy dispensing is considered a more reliable source of information on drug use than patient reporting[11] or physician prescribing, as 20–60% of initial prescriptions are not filled.[12] Although neither of these features alone is sufficient for drawing robust conclusions, meaningful combinations of codes can substantially improve the measurement of events. Having three diagnoses of diabetes within 6 weeks plus the dispensing of an oral antidiabetic medication make it more likely that a patient has diabetes than a single diagnostic code would.[13]

Electronic health records (EHRs) increasingly help augment such claims databases (**Table 1**). EHR systems may contain laboratory test results, imaging test results, biomarker status, and genomics information. However, because EHR systems usually do neither to cover the full population nor the entire care continuum, better methods are needed to fill in the many "blanks," for example, by calibrating claims data information using a subset of patients for whom we have richer information.[14,15]

Analyze suitable data locally and aggregate insights globally

If investigators require more detailed information than is captured through routine health care operations in their local system, they may link several local sources to add the missing details or search elsewhere for the most suitable data source, sometimes outside their jurisdiction.

It is critically important to work closely with local experts who know the exact interpretation of "their" data based on a deep understanding of the specific data generation mechanism. Despite attempts to catalog claims and EHR databases in multiple countries,[16,17] much information about the existence, quality, and accessibility of databases is dynamic and often orally shared among local expert users. Ongoing efforts to improve the information available about databases internationally should provide a

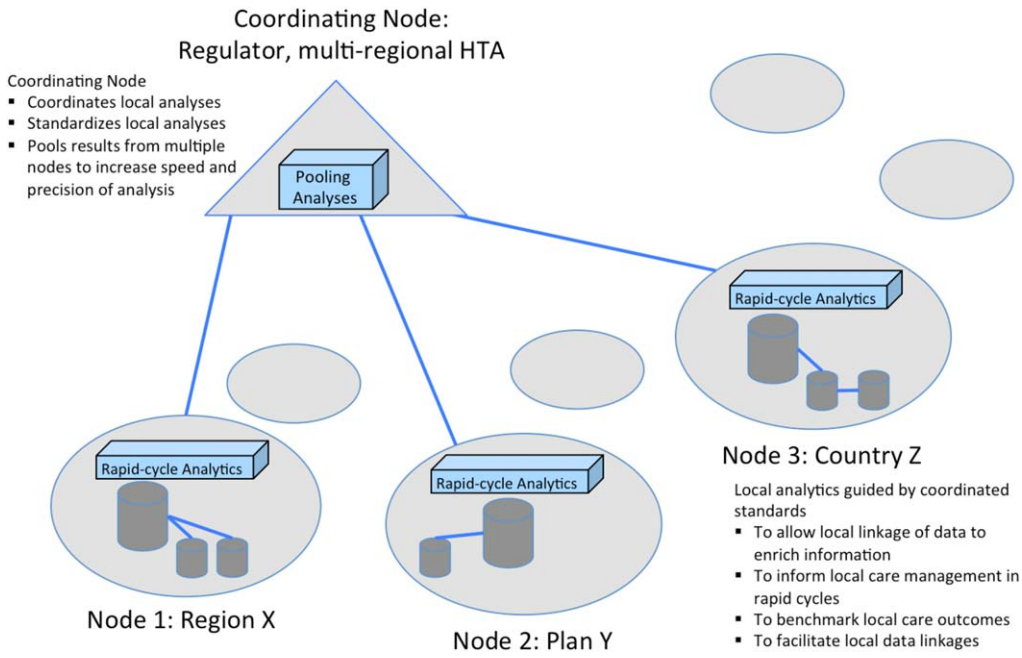


Figure 1 A flexible network of analytic nodes will support global information needs and comply with local privacy and governance rules. *Each analytic node first satisfies local evidence needs for care management and secondarily can be used for rapid-cycle analytics as part of a network. Local and regional governance models will be implemented for each node.

broader capability to combine, contrast, and compare research results.

Many data generators (providers, payers, and registries) today insist that their data remain locally stored, citing data security and patient privacy concerns; only aggregate-level results may leave their systems after local analysis of individual-level data.[18] Linkage of additional data sources to enrich clinical information faces fewer privacy hurdles when conducted locally by the data generators. This leads logically to a network of multiple analytic nodes, each of which may be the result of local linkage activities itself (**Figure 1**). The US Sentinel System, Exploring and Understanding Adverse Drug Reactions, and European Medical Informatics Framework follow this principle.[19,20] The data remain at the site of the data generators, including all local security and privacy precautions, and mostly aggregated results will be shared in the network.[21] New privacy regulations, for instance, in Europe, directive 2016/680, may make the organization of such networks more complicated.

Database flexibility requires us to work with data that vary in terms of quality, content, and coding; this may not be terribly satisfying for researchers used to working with *de novo* generated (primary) data collected for specific research purposes. Learning to cope with data-quality issues is the price we pay for understanding how medications work outside highly controlled research environments and getting a candid picture of utilization patterns in routine care without a researcher's interference. Thus, database researchers frequently choose to work with clever proxy definitions of the underlying health constructs in terms of outcomes[22] or patient characteristics for confounding adjustment.[23] Such proxies are often specifically linked to the intricacies of each health care system, and how it covers and records health services.

The quantity and completeness of data recorded for an individual depends on the frequency of health care encounters. A chronically ill patient has many opportunities to have his/her health status assessed and recorded and is less likely to move and change health plans than healthy adults, leading to voluminous and longer-term data collection, but also to possibly unintended selection of patients for whom data are most complete. In addition, highly integrated health care systems and national systems often have more complete information than typical US commercial insurances.

Tracking use in a target population

Part of an adaptive strategy for innovation is to characterize, target, and track the population who should and actually will receive a new drug. Health care administrative data can track the dispensing of outpatient pharmacies and i.v. administration of specialty medicines almost in real time and are quite valuable in tracking both intended and unexpected uses reliably and expeditiously. This lays the foundation for analyses stratified by specific indications (whether approved or not) and monitoring a variety of usage patterns, including switching, discontinuations, and suboptimal adherence.

Treatment effect measures that support decision-making

To fully evaluate the net value of new medications, decision-makers consider multiple endpoints of benefit and harm simultaneously. A net benefit is established by ascribing preference or quality of life weights to absolute treatment effect measures of

intended and unintended effects.[24] It is therefore imperative that RWD analysis produce measures of the absolute effect size, including risk differences, rate differences, or differences of mean costs. Under an additive causal model, which is most frequently assumed,[25,26] difference measures do not vary with baseline risk,[27] and have population denominators, which make them immediately useful for care management. Epidemiologists underutilize risk-adjusted difference measures, although propensity score matching[28] and regression models enable such estimation for dichotomous and continuous outcomes.[29,30]

It makes sense to provide treatment effect measures for each outcome separately; this allows decision-makers to assign weights according to population preferences and quality of life measures that specifically reflect their constituency. Net benefits can then be computed based on customized assumptions.

Rapid cycle evaluation of economic outcomes can be equally important and actionable for many stakeholders in the health care system, particularly if the budget impact is large and significant value judgments must be made. Most of the same considerations around data completeness, linkages, and treatment effect estimation enumerated here apply to economic outcomes as well, although sometimes with specific nuances depending on the variables and populations involved.[31]

Valid evidence

To meaningfully inform regulators and payers, findings need to be of sufficiently high validity to allow causal conclusions with confidence. In medicine, noncausal associations are insufficient to justify providing access to innovations.

Identifying the most appropriate choice of study design and analytic strategy, in search for the best possible approximation of a counterfactual experience in patients had they not received the innovator drug, does not only require substantial expertise but unavoidably comes with tradeoffs; a perfect solution to evidence generation is rarely possible in a single study regardless as to whether it is a database analysis or other. It is frequently necessary to combine different methodological approaches to the same question (e.g., pairing baseline randomization with secondary data and observational analyses). This dilemma is reflected in guideline documents that share a nonprescriptive attitude toward design and analytic choices.[4,5,7,10] Structured tools for selecting optimal study designs also help in the selection process and make design choices and tradeoffs more transparent.[32]

Study design options for health care data analysis of comparative effects

There are three fundamental types of study designs that can be used to compare the effect of medical products using health care databases: (1) parallel-group cohort designs; (2) self-controlled designs; and (3) cluster designs for group-level comparisons.[33]

The cohort design is the most versatile and by far the most commonly used. New users of the newly marketed drug are compared to new users of a clinically reasonable comparator drug.[34] The groups are compared in terms of event risk after medication start during the same calendar time, so that health system changes or changes in recording practices are inherently taken into consideration (**Figure 2**). If additional information like biomarkers or genomics data are collected and linked to a database study, sampling plans like case-control, case-cohort, or two-stage sampling can be implemented to increase statistical efficiency. In cohorts, it is straightforward to estimate outcome rates, resource use, and costs over defined time periods and to estimate absolute effect sizes (risk differences). The main concern of the cohort design is whether a balance of relevant risk factors between the two comparison groups can be achieved without the help of baseline randomization. This is a case-by-case decision that requires expert assessment, knowledge of the underlying database, and some diagnostics.[35]

Self-controlled designs take advantage of the fact that patients initially unexposed to an agent then become exposed, and their health status may change accordingly. Such designs are indexed either by an event (case-crossover design)[36] or by exposure (pre-post design or self-controlled case series; see **Figure 2**).[37] Because self-controlled designs compare health outcomes within the same patient between an exposed and an unexposed time period, they inherently control for time-invariant patient characteristics; this is a substantial advantage in database studies, in which concerns persist that not all confounding factors can be measured fully. However, self-controlled designs are somewhat susceptible to bias introduced by externalities that may modify the risk of health outcomes independent of the study exposure. Therefore, this design is limited to establish the effectiveness of newly marketed medications when the underlying condition is stable or predictably worsening over time without treatment changes and there is little variation in this decline among patients. The design is also suitable if the new drug almost completely replaces existing medications in a very short time, or if the onset of the new drug's effect is immediate and the effect size substantial.[38] Such observational versions of self-controlled designs can be enhanced by randomization through a stepped-wedge design, which will mitigate any self-selection of patients by rate of declining health.[39]

Tradeoffs to be considered when making analysis choices

Any design and analytic choice comes with tradeoffs. The larger process of choosing the appropriate evidence-generation mechanism can be subdivided into consideration of the study design, the data source, and the analytic strategy (**Figure 3**). The study designs described above can be analyzed in many different ways. Most analytic choices are tied to methods that have been used for confounding control and modeling follow-up.[40]

The primary goal of confounding adjustment is to achieve balance of observed baseline characteristics between treatment groups. In health care databases, the information collected is not under the control of the investigator, therefore, the constructs of interest may be measured in indirect ways using combinations of observable events in various coding systems. Propensity score methods have established themselves as the preferred methodology in database analyses because they can reduce a large number of potential confounders and proxies thereof into a single score variable improving estimation robustness.[23,41]

The decision on the follow-up model is tightly linked to the underlying biology and proposed drug effect. For short-acting

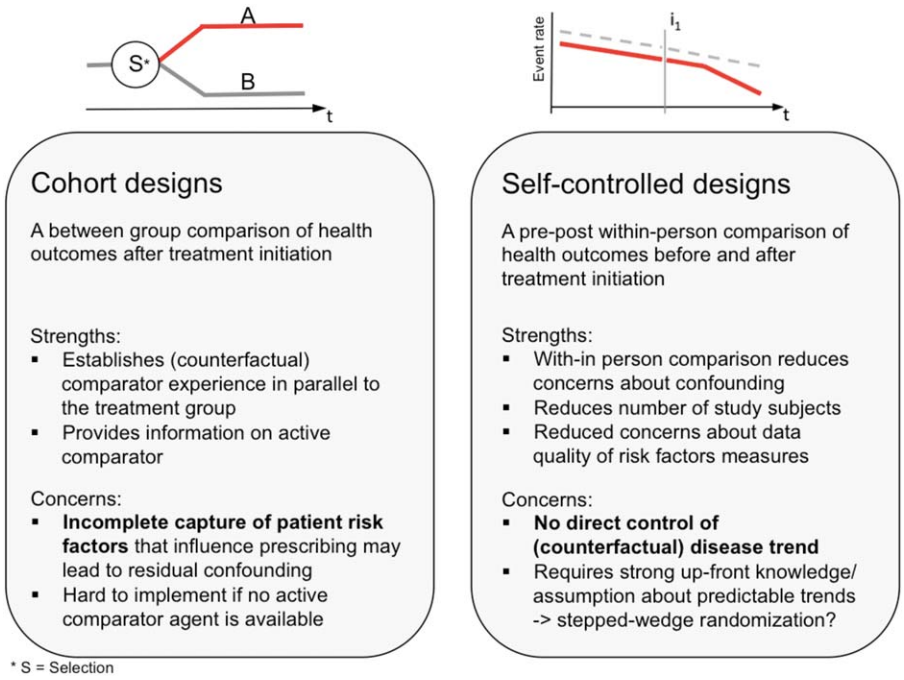


Figure 2 Two frequently used design classes for non-randomized evidence generation with health care databases.

oral medications an as-treated model may be most appropriate, whereas longer-term effectiveness may be better reflected in first-exposure-carried-forward models, which is similar to an intention-to-treat analysis and reduces the concern about informative medication stopping.[42] Other therapies may have delayed onsets and such a lag time needs to be modeled accordingly.[43]

As the number of choices grows, it is difficult to give detailed recommendations for design and analytic approaches that do not result in a textbook. The purpose of **Figure 3** is instead to encourage investigators to be clear about their evidence needs, be fully transparent with their research design choices to satisfy such evidence needs, be aware of the tradeoffs they are making, and justify them in light of the research question and stakeholders involved.

A few dimensions drive study design choices with RWD

One can readily identify four dimensions that drive study design choices and data needs for ABI (**Figure 4**), although additional considerations may arise. One may wish to principally establish existing evidence on the effectiveness and safety of the initial indication or supplement it; if the goal is to test the claim of a new secondary indication, this may require a different level of evidence.

The nature of the treated condition (e.g., whether it is a chronic disease with monotonically declining health or an acute episodic condition), may determine the approach to evidence generation. Historically, controlled studies and prepost self-controlled analyses can be valid choices in evaluating highly targeted therapies with well-characterized molecular mechanisms for treating conditions that would otherwise result in a steady declining health; such designs have been permitted by regulators to establish effectiveness in these contexts.[44] A strictly observational prepost study can be enhanced by stepped-wedge designs that introduce a baseline randomization component delaying treatment in some patients.[45]

The rapid vs. delayed onset of a therapeutic effect and its expected effect size may determine the choice of study design. Treatments that promise substantial effect sizes that materialize quickly may be evaluated with self-controlled prepost designs. Interrupted time-series analyses are viable options when new breakthrough medications are quickly and nearly universally adopted by the provider community, leaving almost no patients on comparator treatment.

Last, multiple stakeholders may focus on different aspects of effectiveness and safety, although all wish to measure those effects with the highest possible validity. Some Health Technology Assessment institutions desire measuring quality of life endpoints or other patient-reported outcomes, which has implications for choosing data sources with direct reach-through to patients.

The dimensions outlined above produce a matrix of preferred design choices for decision-makers to consider; **Table 2** illustrates the numerous possible combinations of study characteristics. This table is further stratified according to the purpose of the generated evidence (i.e., whether to demonstrate effectiveness, harm, utilization, or value; cost per net benefit). This additional stratification is necessary because data needs may differ substantially. For example, claims data are particularly strong in monitoring utilization and assessing cost to a care system under routine circumstances; for harm investigations, the key question is whether RWD can capture the outcomes of concern; and for effectiveness assessment, the central consideration is whether sufficient

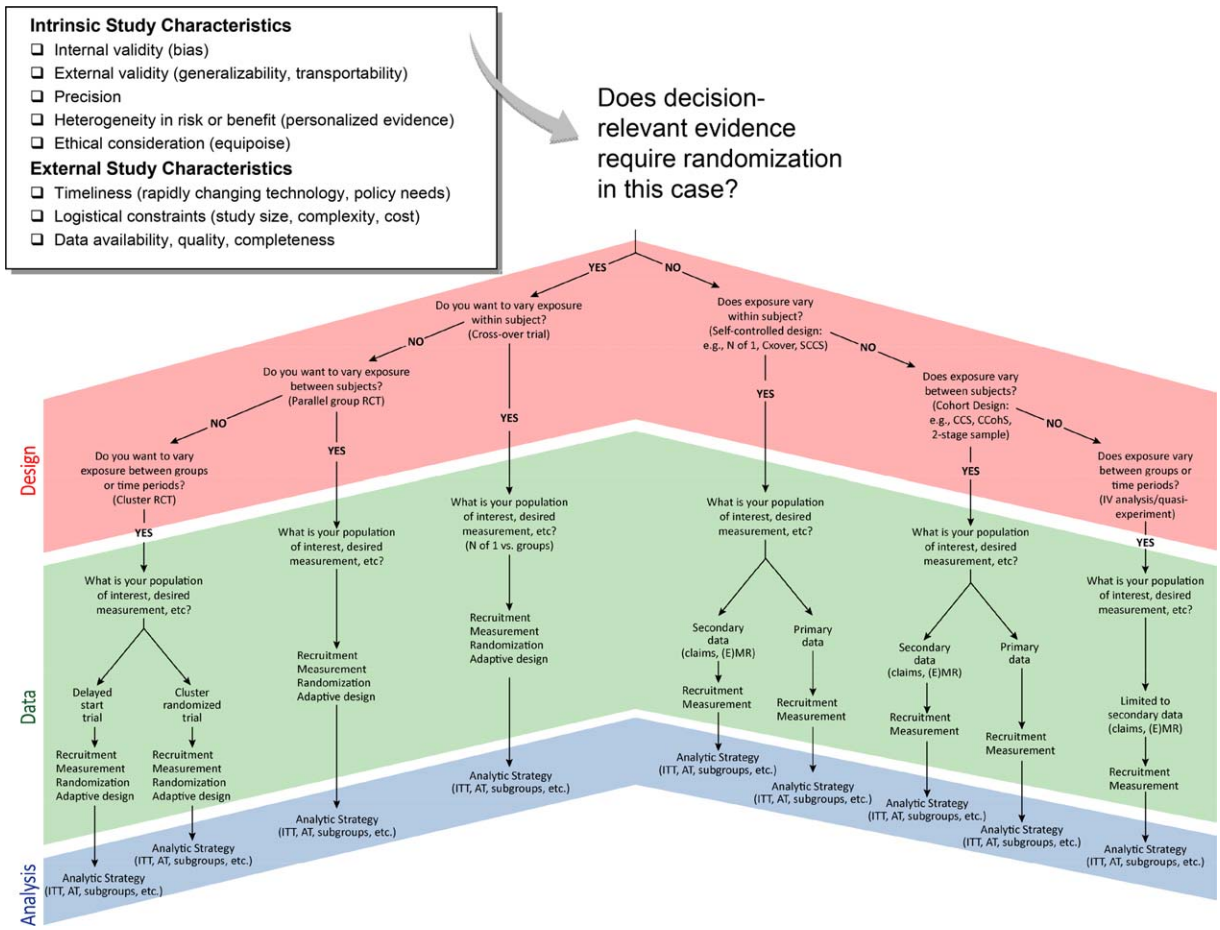


Figure 3 An overview of study design choices structured by design, data, and analysis choices. *Although this tree characterizes the principles of choices and tradeoffs it represents only a limited set of choices that are frequently made but is by no means comprehensive.

confounding control can be achieved without randomization. The supplementary materials include four worked examples (**Supplementary Tables S3–S6** online).

Expedited evidence from dynamic data sources

In an ABI process, the clock is constantly ticking; decision-ready evidence needs to be developed and updated rapidly to ensure that patients are treated with effective and safe drugs based on the best possible knowledge at a given time point. The incremental development of evidence requires expeditious analyses of constantly refreshing dynamic data sources generated by health care systems.[46] Three key factors that influence the speed with which evidence can be generated from RWD like health care databases are (1) time between data generation and data access; (2) time elapsed between data access and completed analysis; and (3) the size of the combined source population in which new users of the target drug are identified.

Shortening time from data generation to data access

Electronic health care data are generated almost instantaneously as care is provided. However, typical health care data sources are made available for analysis only after substantial lag time. Most commercial databases lag 6 to 9 months behind until data become available to researchers. Lag time can be up to 3 years, an unacceptable delay for meaningful support of iterative decision-making with RWD.[47] There are, however, a number of health care systems that aggregate a constant stream of the most relevant health care data in almost real time. Such systems with real-time data access may not want to send their data to external parties for analysis and may have limited interest in participating in multiregional analysis of the effectiveness of new drugs. In order to actively involve them in ABI, they can be supported with local analytic nodes to satisfy local evidence needs for value-based care management, but secondarily can be used for rapid-cycle analytics on drug effects as part of an analytics network (**Figure 1**). An added advantage is that linkage to relevant data sources (e.g., vital statistics, socioeconomic indicators, or behavioral data), can be achieved more easily on the level of the data generator than on a higher aggregated level. Local analyses that follow a standardized implementation protocol will yield findings that will be pooled across multiple participating nodes, avoiding unnecessary wait-times that may arise if data refresh asynchronously.

Interjecting the data aggregation with high-level common data models may slow the process as new mapping needs to be completed each time data refresh, may lead to misrepresentations of the data,[48,49] and may ultimately be misconstrued as data standardization, although like-looking data constructs may be based on quite different information content (e.g., a myocardial infarction may be

		Examples	Implications for **effectiveness** evidence needs
1 Purpose	A) **Confirming** initial indication	▪ Biomarker endpoint needs confirmation using clinical endpoints ▪ Unclear RCT findings need confirmation ▪ Evidence from highly ctrl'ed research needs confirmation in routine care settings	Fairly stringent evidence needs of effectiveness and safety
	B) **Broadening** indication	▪ Different cancer but same molecular mechanism leads to off-label use: Effectiveness in off label? ▪ Initial treatment in severe patients expands to less severe (early stage): Effectiveness in new patient subgroups?	Borrow safety information from primary indication Observational effectiveness?
2 Condition	A) **Monotonic** disease progression	▪ Predictable constant decline of health e.g. many genetic conditions	Observational pre-post design possible Consider step-wedge design
	B) **Episodic**, undulating disease progression	▪ Unpredictable variation in disease severity over time, e.g. bipolar affective disorder	IIb RCT +/- cohort
3 Therapy	A) **Rapid onset of effect** and substantial effect size	▪ Obvious clinical improvement shortly after therapy start e.g. gene-therapy, enzyme subst.	Consider pre-post design, step-wedge design
	B) **Incremental** effectiveness	▪ Medium-targeted cancer treatment ▪ Innovative oral antidiabetic drug	III RCT
4 Stake-holder	A) **Regulatory agencies**	EMA providing marketing authorization under MAPP, focused on efficacy vs. placebo or vs. active	New policies and exemption rules provide pathways for innovate analyses
	B) **Payers, HTA agencies**	Focused on incremental value, i.e. incremental net benefit	Additionally requires patient reported outcomes outcomes

Figure 4 Dimensions driving study design choices and data needs for incremental evidence generation as part of adaptive biomedical innovations.

recorded based on a single diagnostic code or on troponin laboratory test results). Common data models do not solve any underlying discrepancy in information content between data sources.[50]

Shortening analysis cycle time

Shortening analysis time when working with health care databases is today no longer a question of computing power. Overall analysis time is determined more by how data are prepared and aggregated, how efficiently complex longitudinal queries can be implemented and run in a reliable way, how fast study designs and sensitivity analyses can be implemented, how statistical analyses can be parallelized and automated, and how comprehensively an automated reporting mechanism informs its users.[51] Once analysis cycles can be shortened without compromising analytic quality, we can move beyond traditional one-off analyses (**Figure 5a**) toward a system that dynamically monitors developing evidence on the effectiveness, safety, and utilization of newly marketed drugs.[52] In such systems, analyses can be repeated each time the data stream refreshes in a local data warehouse, or whenever the number of new drug users grows by a fixed increment that is relevant to decision-makers (**Figure 5b**).[53]

Such a rapid-cycle monitoring system has distinct advantages. Precision of a given point estimate for a drug-outcome relationship becomes slightly less relevant, because precision improves as more patients use the new drug, and to some degree precision gains can be extrapolated based on observed trends. There is some uncertainty among decision-makers of how to deal with multiple comparisons that arise from such a system. Because no health care intervention/program takes place in a hypothesis-free space, typical approaches like *P* values with corresponding adjustments for multiple testing that may reject true benefits,[54,55] and shrinkage of effect estimates toward the null, are of little value to decision-makers. Decision-makers are more likely focused on benefit-harm considerations that are driven by the absolute effect size (difference measures) across several weighted endpoints.

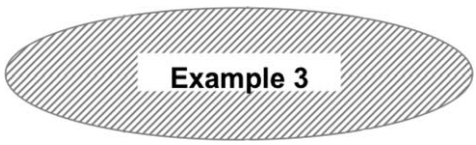
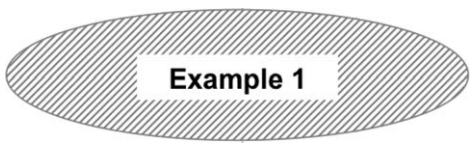
Increasing source population

RWD on newly marketed medications from health care databases always suffers from the "catch 22" that if providers adopt a new drug slowly, few data on safety and effectiveness are generated by health care databases, which in turn will slow its use. A reasonable approach to mitigate this problem is to dramatically increase the source population that is monitored. Even if the proportion of target patients receiving the new drug is small, the absolute number of users can be increased for evaluation purposes, which will improve the precision of the estimated treatment effectiveness in early interim analyses.

A flexible multidatabase system with decentralized analytics can be scaled to fit the evidence needs for a new medical product of varying market penetration. Regional differences in effectiveness can be identified (**Figure 5c**), and pooled estimates across regions will stabilize more quickly with higher precision (**Figure 5d**), optimally supporting incremental decision-making. Models for combining results from multiple database analyses[19,56–58] are capable of analyzing population sizes exceeding 50 million at a given point in time. However, they require comprehensive governance models that accommodate different local regulations for data access.

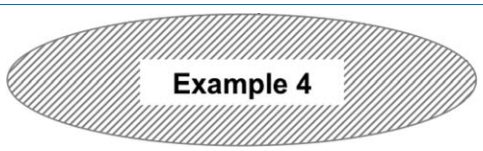
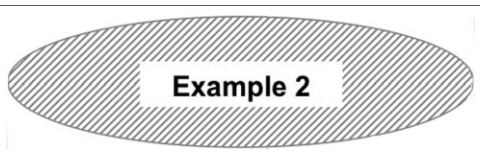
Table 2 Context-driven study design choice matrix

A. Confirmation of initial approval: Design options and RWD

Initial approval	Likely INCREMENTAL improvement				Likely SUBSTANTIAL improvement			
	Effectiveness intended trt. effects	Harm Un-intended trt. effects	Utilization	Value[a]	Effectiveness intended trt. effects	Harm Un-intended trt. effects	Utilization	Value[a]
Small molecule or biologic with biomarker endpoint								
Rare disease, e.g. gene therapy, enzyme replacement therapy								
Targeted indication, e.g. expression-based anti-cancer med'n, anti-cancer immunotherapy								

[a] Cost per net benefit

B. Expansion of initial indication: Design options and real world data (RWD)[a]

Initial approval	Likely INCREMENTAL improvement				Likely SUBSTANTIAL improvement			
	Effectiveness intended trt. effects	Harm Un-intended trt. effects	Utilization	Value	Effectiveness intended trt. effects	Harm Un-intended trt. effects	Utilization	Value[a]
Small molecule or biologic with biomarker endpoint								
Rare disease, e.g. gene therapy, enzyme replacement therapy								
Targeted indication, e.g. expression-based anti-cancer medication, anti-cancer immunotherapy								

Study design and data source considerations are influenced by several dimensions leading to a matrix of design options: Initial approval vs. expansion of indication; incremental vs. substantial effect; type of study purpose (i.e., whether to demonstrate effectiveness, safety, utilization, or value). As individual choices are highly context-dependent, the design choices are illustrated in four worked examples in the **Supplementary Tables S3–S6** online.

Transparent and reproducible evidence

Ultimately, big data analyses should inform evidence-based decisions by people who usually are not the ones generating the evidence. Typically, when the decision-maker is not the evidence generator, there would need to be tremendous trust in the way the relevant evidence was produced. However, because trust is neither objectively measured nor scalable, it needs to be replaced by good governance in the production and handling of the relevant evidence so that there can be confidence in the processes and outputs. Such an approach starts with maintaining the highest possible transparency and accountability in conducting database studies, which will allow reproducing and verification of

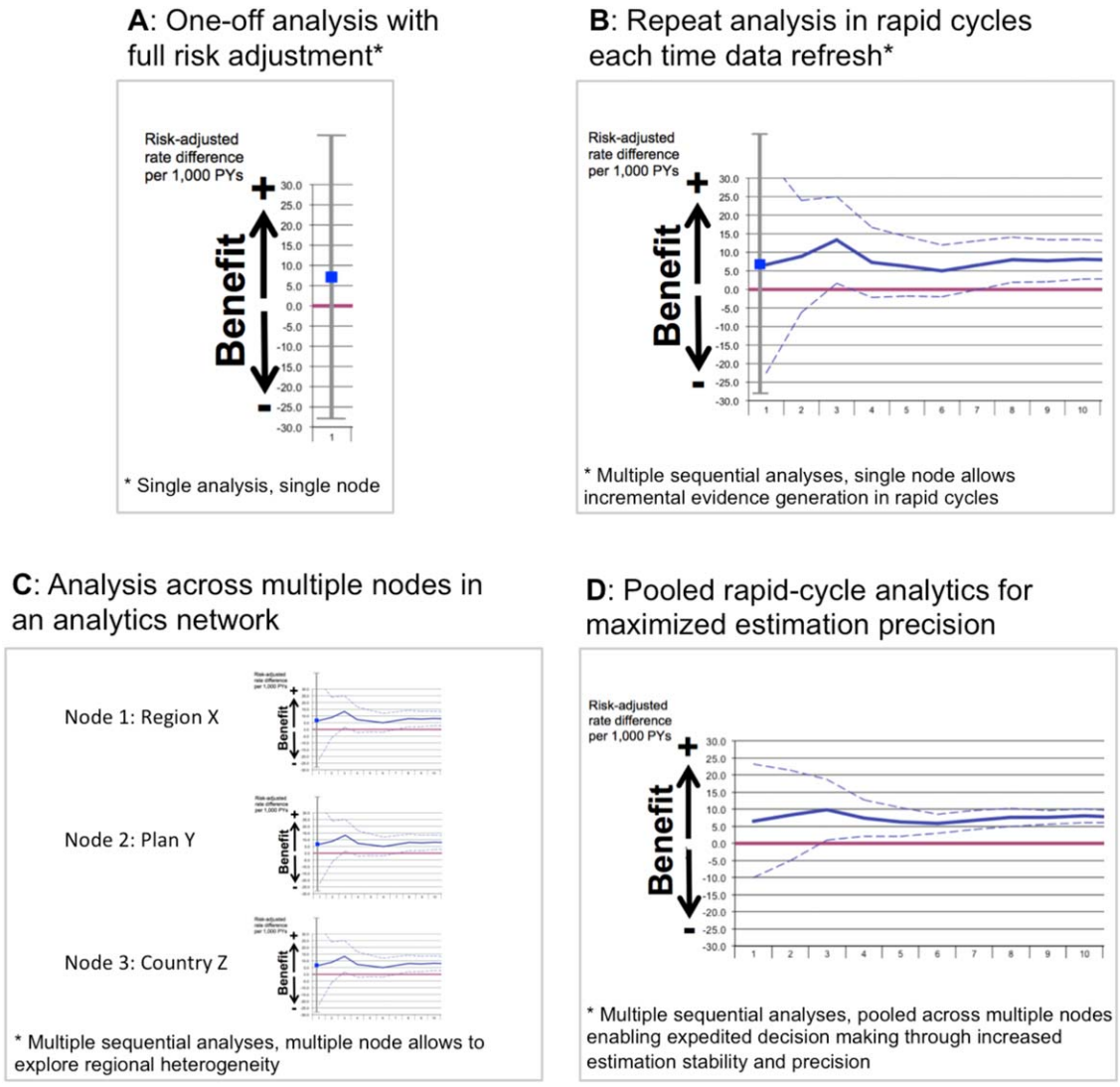


Figure 5 Expedited evidence generation through increasing frequency of analyses and larger size of source population. *Synthetic data for illustration only.

results using the same data source. Reproducibility is, in turn, the foundation for replication and robustness checking to support causal conclusions.[59]

Transparency

Confidence in database analytics is grounded in transparency, which makes it possible to reproduce results using the same data. Decision-makers lose confidence in findings if reproduction of the results is not possible due to the lack of transparency or because one tried and failed to reproduce. Researchers engaging in primary data collection have developed several vehicles to deposit original research data to allow reproduction outside a regulatory environment and journal editors have decided to require data deposition for randomized controlled trials alongside a publication. However, when working with secondary data that were not collected for research purposes and were not curated, many decisions must be made when implementing a specific study. If the specifics of those choices are not explicitly stated, a researcher trying to reproduce a study will obtain numerically different results even when working with the same source data. Let us take as an example this question: Does follow-up begin on the day of first exposure or the following day? This might seem irrelevant but could have a substantial impact on study findings, including qualitatively different conclusions. A spurious association between benzodiazepines and gastrointestinal bleeds was reported because the first benzo exposure was allowed to occur the same day as the gastrointestinal bleed outcome.[60] Although this could have been an immediate, dramatic adverse event caused by benzos, it is much more likely that benzos were simply used in the treatment of the gastrointestinal bleed.[61]

Recent work showed that many published database studies still have suboptimal transparency, largely because sharing the details of study implementation and coding is insufficient to transform raw data into meaningful variables.[8] An infrastructure in which

to deposit study protocols or code lists is being developed for database studies, including the European Network of Centres for Pharmacoepidemiology and Pharmacovigilance,[62,63] and its use is encouraged by decision-makers. The requirement of maximum transparency is independent of whether study details are predefined and protocols registered/deposited, or whether full disclosure of the presented analysis and all precursor analyses happens once a study is completed. The RECORD guideline is a starting point to encourage clearer reporting of database studies. However, more prescriptive guidance on the reporting of technical details may be necessary to enable reproduction. The US Food and Drug Administration Sentinel program is implicitly establishing best reporting practices by making public all key parameters necessary to run any of their preprogrammed software tools to implement a study.[32,64]

Replicability

Once reproducibility is achieved, either through fully transparent reporting or by actually reproducing a study, a causal biologic relationship is more likely if the findings can be replicated in different populations. This may be in the same geographic region or in different health care systems. A key issue for replication of evidence is that, in secondary database studies, the information content may vary (sometimes substantially) even if the same codes are used to define medical constructs. A study conducted using Danish claims databases identified a meaningfully lower prevalence of cardiovascular conditions than a comparable study with US claims data; the Danish database did not include information from outpatient visits,[65,66] which lowered the number of opportunities for the electronic health care data stream to record the existing conditions of study subjects. To achieve the highest validity with a given body of data, it is advisable to let the researchers who are most familiar with the local data-generation mechanisms —and therefore the data quality—design and implement the optimal coding.

Robustness

Quite often, there are disagreements among decision-makers or researchers and reviewers regarding design assumptions and analytics choices in implementing a study. In reproducible and replicable studies, which is enabled by transparency, one can then test how robust the reported findings are toward slightly different design choices (e.g., differences in the included patient population,[67] variations in follow-up model, inclusion of new users versus a mix of existing and new users,[65] and many other choices).[33] The US Food and Drug Administration Sentinel program implemented a systematic assessment of sensitivity analysis in database studies and prioritized possible activities.[64]

Such sensitivity analyses of existing studies are powerful tools for decision-makers who will feel much more confident if the qualitative interpretation of a study does not change after assumptions are slightly altered.[33,68] Currently, sensitivity analyses involve cumbersome reanalyses of data or collaboration with the original investigator team to change their analysis. Such robustness checks should be built into the workflow of newer software to avoid those labor-intensive processes.

External constraints

The generation and analysis of RWD takes place in an increasingly regulated environment. Adaptive innovations may disrupt older, accustomed pathways, and local, regional, and national system architectures, processes, responsibilities, and regulations may not yet be developed or coordinated accordingly.

Among the issues raised by regional health plans are the perceived lack of a legal, organizational, and financial framework to facilitate adaptive evidence generation and regulatory decision-making. Specifically, this includes unresolved questions about local data quality and enforcing data quality standards, more stringent requirements regarding transparency and demonstrated replicability and robustness checks, legal means to ensure that data collection and evidence generation are completed in a predefined time frame, and a legal framework to dynamically adjust product prices according to growing evidence of incremental benefit or the lack thereof. The legal authority of a regulatory agency in charge for marketing approval may not overlap with the legal framework of the payer organizations that provide coverage, which leads to uncertainty in the planning of and decision-making for ABIs.[69] In order for the evidence to serve both regulators and payers, the logical orchestration and temporal arrangement of multiple studies needs coordination. Such coordination is increasingly a focus of activity.

A consequence of expedited evidence development through rapid-cycle analytics is that decision-makers have many more opportunities to review incremental findings to inform their decisions. In a paradoxical way, this may cause hesitations. Even very important insights may initially be difficult to spot when uncertainty is large and confidence intervals wide. Furthermore, it is not resource neutral for agencies to review all incoming findings. In the most extreme, this could theoretically result in a wish to review only the final study results after a protracted period of study time. Clearly, requirements for continuous monitoring of emerging evidence must accord with patients' justified demand that the responsible agencies closely monitor the effectiveness and harm of the products they use. In addition to obvious resolutions of the staffing need, this issue can be remedied by providing highly transparent and semistandardized reports. The assessor needs to gain a high level of confidence in an analysis faster than usual without compromising the quality of the review. This may be facilitated by using validated analysis tools and highly structured reporting that eases reading and does encourage fullest transparency.

For many plans, it is difficult to incorporate new evidence into existing benefits plans outside of a set calendar that is agreed with external parties. In addition, all changes to a plan result in member notifications. Together, this results in a discrepancy between the capability to incrementally develop evidence as data refresh and the effective opportunities for changing past decisions.

These concerns regarding iterative evidence development on the part of health system decision-makers need thoughtful consideration in order for an ABI system to work most effectively.

CONCLUSION AND OUTLOOK

Electronic health care data streams are ubiquitous in any modern health care system, offering plentiful sources of meaningful

information to decision-makers committed to adaptive innovation. Two key characteristics make such data sources uniquely valuable: (1) they reflect how medical products impact health in a target population under routine care outside a highly controlled research environment; and (2) the constant stream of data, which can be analyzed in rapid cycles, enables incremental evidence development for real-time capture of benefit-harm profiles. In a single payer/provider organization with their longitudinal data assets, evidence generation can be directly linked to care management; further, evidence may be pooled across many entities to achieve faster reduction of clinical/value uncertainty typically associated with new drugs and enable more timely, evidence-based decisions on how to use innovations most effectively. These aspects are central to an adaptive innovation paradigm.

Two issues that demand continued attention are the quality of data to answer a specific question and the apparent variation in the validity of findings. Unfortunately, these parameters are highly dependent upon context, which has made it difficult for the field to produce prescriptive best-practice statements comparable to those already accepted for randomized controlled trials. As shown in the worked examples above, for one therapy, a database study might be appropriate to demonstrate effectiveness and value, but for another product, it may be suitable to monitor materialized harm.

The MVET framework presented above is helpful in structuring a discussion on how to most effectively use database analytics to create an adaptive innovation environment and to confirm that key requirements for database evidence are of maximum utility for decision-making, whether regulator, payer, provider, or product sponsor. Over time, the use of a framework like MVET will enable the community to converge on standards for generating evidence with health care database studies to inform value-based care management in target populations.

Given their level of context-dependency, some aspects of the framework (transparency, expedited evidence) have the potential to develop into normative standards earlier than others (validity and data quality) through accumulation of use cases and adaptation of existing guidelines. Simultaneously, stakeholders can begin to facilitate a process by removing external obstacles (i.e., adjusting rules and regulations), recognizing that some will take longer than others.

Rapid-cycle analytics of health care databases is maturing as acceleration of value-based care drives demand for such evidence. Governance, regulations, and data quality are catching up as the utility of this resource is demonstrated in multiple contexts.

CONFLICT OF INTEREST

S.S. is principal investigator of research contracts from Genentech and Boehringer Ingelheim to the Brigham and Women's Hospital from which he receives salary support; he is a consultant to WHISCON LLC and Aetion Inc., of which he holds equity. C.C. is employee of Sanofi, a pharmaceutical manufacturer. A.-V.E. is an employee of Bluebird Bio, a biotechnology manufacturer. S.P. is an employee of Humana. M.R. is an employee of Aetion Inc., a software manufacturer. B.V. is an employee of Janssen, a pharmaceutical manufacturer. R.J.W. is an employee of the International Society of Pharmacoeconomics and Outcomes Research. H.-G.E., A.G.-A., S.G., W.G., R.L., W.L., D.M., T.M., B.J.P., R.P., and A.S. have no conflicts to report.

Additional Supporting Information may be found in the online version of this article.

ACKNOWLEDGMENTS

The views expressed in this article are the personal views of the authors and may not be understood or quoted as being made on behalf of or reflecting the position of the agencies or organizations with which the authors are affiliated. Investigator-specific disclosures are listed at the end of the manuscript.

SOURCE OF FUNDING

Dr. Schneeweiss' research that contributed to this work is funded by grants and contracts from the Patient Center Outcomes Research Institute, the National Institute of Health, and the US Food and Drug Administration. Parts of this manuscript have been presented earlier at the annual meeting of the International Society for Pharmacoepidemiology, and a workshop funded by IMI Get Real, IMI ADAPT SMART, and NEWDIGS.

1. Kesselheim, A.S. & Darrow, J.J. FDA designations for therapeutics and their impact on drug development and regulatory review outcomes. *Clin. Pharmacol. Ther.* **97**, 29–36 (2015).
2. Eichler, H.G. *et al*. Adaptive licensing: taking the next step in the evolution of drug approval. *Clin. Pharmacol. Ther.* **91**, 426–437 (2012).
3. 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).
4. The International Society of Pharmacoepidemiology. Guidelines for good pharmacoepidemiology practices (GPP). *Pharmacoepidemiol. Drug Saf.* **17**, 200–208 (2008).
5. Federal Register. Guidance for industry and Food and Drug Administration staff on best practices for conducting and reporting pharmacoepidemiologic safety studies using electronic healthcare data; availability. <https://www.federalregister.gov/documents/2013/05/14/2013-11361/guidance-for-industry-and-food-and-drug-administration-staff-on-best-practices-for-conducting-and> (2013).
6. European Network of Centres for Pharmacepidemiology and Pharmacovigilance. ENCePP guide on methodological standards in pharmacoepidemiology. <http://www.encepp.eu/standards_and_guidances/methodologicalGuide.shtml> (2014).
7. Johnson, M.L., Crown, W., Martin, B.C., Dormuth, C.R. & Siebert, U. Good research practices for comparative effectiveness research: analytic methods to improve causal inference from nonrandomized studies of treatment effects using secondary data sources: the ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report–Part III. *Value Health* **12**, 1062–1073 (2009).
8. Wang, S.V., Verpillat, P., Rassen, J.A., Patrick, A., Garry, E.M. & Bartels, D.B. Transparency and reproducibility of observational cohort studies using large healthcare databases. *Clin. Pharmacol. Ther.* **99**, 325–332 (2016).
9. Schneeweiss, S. Improving therapeutic effectiveness and safety through big healthcare data. *Clin. Pharmacol. Ther.* **99**, 262–265 (2016).
10. European Medicines Agency. Scientific guidance on post-authorisation efficacy studies. <http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2015/11/WC500196379.pdf> (2015).
11. West, S.L., Savitz, D.A., Koch, G., Strom, B.L., Guess, H.A. & Hartzema, A. Recall accuracy for prescription medications: self-report compared with database information. *Am. J. Epidemiol.* **142**, 1103–1112 (1995).
12. Fischer, M.A. *et al*. Primary medication non-adherence: analysis of 195,930 electronic prescriptions. *J. Gen. Intern. Med.* **25**, 284–290 (2010).

13. Glynn, R.J., Monane, M., Gurwitz, J.H., Choodnovskiy, I. & Avorn, J. Agreement between drug treatment data and a discharge diagnosis of diabetes mellitus in the elderly. *Am. J. Epidemiol.* **149**, 541–549 (1999).
14. Stürmer, T., Schneeweiss, S., Avorn, J. & Glynn, R.J. Adjusting effect estimates for unmeasured confounding with validation data using propensity score calibration. *Am. J. Epidemiol.* **162**, 279–289 (2005).
15. Schneeweiss, S. *et al*. Supplementing claims data with outpatient laboratory test results to improve confounding adjustment in effectiveness studies of lipid-lowering treatments. *BMC Med. Res. Methodol.* **12**, 180 (2012).
16. B.R.I.D.G.E. to data. The search engine for healthcare databases. <http://www.bridgetodata.org/> (2016). Accessed June 2016.
17. International Society for Pharmacoeconomics and Outcomes Research: ISPOR International Digest of Databases. <http://www.ispor.org/digestofintdb/countrylist.aspx> (2016). Accessed June 2016.
18. Mandl, K.D. & Kohane, I.S. Federalist principles for healthcare data networks. *Nat. Biotechnol.* **33**, 360–363 (2015).
19. Platt, R. *et al*. The U.S. Food and Drug Administration's Mini-Sentinel program: status and direction. *Pharmacoepidemiol. Drug Saf.* **21** Suppl 1, 1–8 (2012).
20. Laheij, R.J., Sturkenboom, M.C., Hassing, R.J., Dieleman, J., Stricker, B.H. & Jansen, J.B. Risk of community-acquired pneumonia and use of gastric acid-suppressive drugs. *JAMA* **292**, 1955–1960 (2004).
21. Curtis, L.H. *et al*. Design considerations, architecture, and use of the Mini-Sentinel distributed data system. *Pharmacoepidemiol. Drug Saf.* **21** Suppl 1, 23–31 (2012).
22. Carnahan, R.M. & Moores, K.G. Mini-Sentinel's systematic reviews of validated methods for identifying health outcomes using administrative and claims data: methods and lessons learned. *Pharmacoepidemiol. Drug Saf.* **21** (suppl. 1), 82–89 (2012).
23. 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).
24. Najafzadeh, M. *et al*. A unified framework for classification of methods for benefit-risk assessment. *Value Health* **18**, 250–259 (2015).
25. Rothman, K.J. & Greenland, S. Causation and causal inference in epidemiology. *Am. J. Public Health* **95** (suppl. 1), S144–S150 (2005).
26. Rubin, D.B. Estimating causal effects from large data sets using propensity scores. *Ann. Intern. Med.* **127**(8 Pt 2), 757–763 (1997).
27. Rothman, K.J. *Modern Epidemiology 3rd Edition* (Philadelphia, PA, Lippincott Williams & Wilkins, 2008).
28. Gagne, J.J., Wang, S.V., Rassen, J.A. & Schneeweiss, S. A modular, prospective, semi-automated drug safety monitoring system for use in a distributed data environment. *Pharmacoepidemiol. Drug Saf.* **23**, 619–627 (2014).
29. Spiegelman, D. & Hertzmark, E. Easy SAS calculations for risk or prevalence ratios and differences. *Am. J. Epidemiol.* **162**, 199–200 (2005).
30. Grootendorst, P.V. A comparison of alternative models of prescription drug utilization. *Health Econ.* **4**, 183–198 (1995).
31. Garrison, L.P. Jr, Neumann, P.J., Erickson, P., Marshall, D. & Mullins, C.D. Using real-world data for coverage and payment decisions: the ISPOR Real-World Data Task Force report. *Value Health* **10**, 326–335 (2007).
32. Gagne, J.J. *et al*. Design considerations in an active medical product safety monitoring system. *Pharmacoepidemiol. Drug Saf.* **21** (suppl. 1), 32–40 (2012).
33. Schneeweiss, S. A basic study design for expedited safety signal evaluation based on electronic healthcare data. *Pharmacoepidemiol. Drug Saf.* **19**, 858–868 (2010).
34. Johnson, E.S. *et al*. The incident user design in comparative effectiveness research. *Pharmacoepidemiol. Drug Saf.* **22**, 1–6 (2013).
35. Brookhart, M.A., Stürmer, T., Glynn, R.J., Rassen, J. & Schneeweiss, S. Confounding control in healthcare database research: challenges and potential approaches. *Med. Care* **48** (suppl. 6), S114–S120 (2010).
36. Maclure, M. The case-crossover design: a method for studying transient effects on the risk of acute events. *Am. J. Epidemiol.* **133**, 144–153 (1991).
37. Farrington, C.P. Relative incidence estimation from case series for vaccine safety evaluation. *Biometrics* **51**, 228–235 (1995).
38. Maclure, M. *et al*. When should case-only designs be used for safety monitoring of medical products? *Pharmacoepidemiol. Drug Saf.* **21** Suppl 1, 50–61 (2012).
39. Shadish, W.R., Cook, T.D. & Campbell, D.T. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference* (Wadsworth Publishing, Bel Air, CA, 2001).
40. Schneeweiss, S. Developments in post-marketing comparative effectiveness research. *Clin. Pharmacol. Ther.* **82**, 143–156 (2007).
41. Rosenbaum, P.R. & Rubin, D.B. The central role of propensity scores in observational studies for causal effects. *Biometrika* **70**, 41–55 (1983).
42. 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).
43. James, S.K. *et al*. Long-term safety and efficacy of drug-eluting versus bare-metal stents in Sweden. *N. Engl. J. Med.* **360**, 1933–1945 (2009).
44. Sasinowski, F.J. *Quantum of Effectiveness Evidence in FDA's Approval of Orphan Drugs* (National Organization for Rare Disorders, Danbury, CT, 2011).
45. Maclure, M., Carleton, B. & Schneeweiss, S. Designed delays versus rigorous pragmatic trials: lower carat gold standards can produce relevant drug evaluations. *Med. Care.* **45** (10 suppl. 2), S44–S49 (2007).
46. Psaty, B.M. & Breckenridge, A.M. Mini-Sentinel and regulatory science–big data rendered fit and functional. *N. Engl. J. Med.* **370**, 2165–2167 (2014).
47. Schneeweiss, S., Shrank, W.H., Ruhl, M. & Maclure, M. Decision-making aligned with rapid-cycle evaluation in health care. *Int. J. Technol. Assess. Health Care* **31**, 214–222 (2015).
48. Zhou, X. *et al*. An evaluation of the THIN database in the OMOP common data model for active drug safety surveillance. *Drug Saf.* **36**, 119–134 (2013).
49. Xu, Y. *et al*. A comparative assessment of observational medical outcomes partnership and Mini-Sentinel common data models and analytics: implications for active drug safety surveillance. *Drug Saf.* **38**, 749–765 (2015).
50. Madigan, D. *et al*. Evaluating the impact of database heterogeneity on observational study results. *Am. J. Epidemiol.* **178**, 645–651 (2013).
51. Schneeweiss, S. Learning from big health care data. *N. Engl. J. Med.* **370**, 2161–2163 (2014).
52. Schneeweiss, S., Gagne, J.J., Glynn, R.J., Ruhl, M. & Rassen, J.A. Assessing the comparative effectiveness of newly marketed medications: methodological challenges and implications for drug development. *Clin. Pharmacol. Ther.* **90**, 777–790 (2011).
53. Stratton, K.G., Cook, A.J., Jackson, L.A. & Nelson, J.C. Simulation study comparing exposure matching with regression adjustment in an observational safety setting with group sequential monitoring. *Stat. Med.* **34**, 1117–1133 (2015).
54. Rothman, K.J. No adjustments are needed for multiple comparisons. *Epidemiology* **1**, 43–46 (1990).
55. Rothman, K.J. Six persistent research misconceptions. *J. Gen. Intern. Med.* **29**, 1060–1064 (2014).
56. Trifirò, G. *et al*. Combining multiple healthcare databases for postmarketing drug and vaccine safety surveillance: why and how? *J. Intern. Med.* **275**, 551–561 (2014).
57. Mandl, K.D. *et al*. Scalable collaborative infrastructure for a learning healthcare system (SCILHS): architecture. *J. Am. Med. Inform. Assoc.* **21**, 615–620 (2014).
58. European Medical Information Framework. <http://www.emif.eu/>. Accessed June 2016.
59. Hill, A.B. The environment and disease: association or causation? *Proc. R. Soc. Med.* **58**, 295–300 (1965).
60. The Observational Medical Outcomes Partnership. <http://www.omop.org> (2015).
61. Gruber, S. *et al*. Design and analysis choices for safety surveillance evaluations need to be tuned to the specifics of the hypothesized drug-outcome association: safety study design and analysis choices need tuning. *Pharmacoepidemiol. Drug Saf.* **25**, 973–981 (2016).

62. European Network of Centres for Pharmacoepidemiology and Pharmacovigilance. <http://www.encepp.eu/>. Accessed June 2016.
63. Springate, D.A. *et al*. ClinicalCodes: an online clinical codes repository to improve the validity and reproducibility of research using electronic medical records. *PLoS One* **9**, e99825 (2014).
64. Mini-Sentinel. <http://www.minisentinel.org/> (2013).
65. Sørensen, R. *et al*. Dabigatran use in Danish atrial fibrillation patients in 2011: a nationwide study. *BMJ Open* **3**, e002758 (2013).
66. Seeger, J.D., Bykov, K., Bartels, D.B., Huybrechts, K., Zint, K. & Schneeweiss, S. Safety and effectiveness of dabigatran and warfarin in routine care of patients with atrial fibrillation. *Thromb. Haemost.* **114**, 1277–1289 (2015).
67. Schneeweiss, S. *et al*. Increasing levels of restriction in pharmacoepidemiologic database studies of elderly and comparison with randomized trial results. *Med. Care* **45**(10 suppl. 2), S131–S142 (2007).
68. Gagne, J.J. *et al*. Mini-Sentinel methods. Analytical methods to assess robustness of drug safety monitoring results. FDA Mini-Sentinel. <http://mini-sentinel.org/work_products/Statistical_Methods/Mini-Sentinel_Methods_Analytical-Methods-to-Assess-Robustness-of-Drug-Safety-Monitoring-Results.pdf> (2015).
69. Schwarzer, R. & Siebert, U. Methods, procedures, and contextual characteristics of health technology assessment and health policy decision making: comparison of health technology assessment agencies in Germany, United Kingdom, France, and Sweden. *Int. J. Technol. Assess. Health Care* **25**, 305–314 (2009).
70. Maclure, M. & Schneeweiss, S. Causation of bias: the episcope. *Epidemiology* **12**, 114–122 (2001).