



eGEMs
Generating Evidence & Methods
to improve patient outcomes

Analytical Methods for a Learning Health System: 1. Framing the Research Question

Michael Stoto, PhD;ⁱ Michael Oakes, PhD;ⁱⁱ Elizabeth Stuart, PhD;ⁱⁱⁱ Lucy Savitz, PhD, MBA;^{iv} Elisa L. Priest, DrPH;^v
Jelena Zurovac, PhD^{vi}

ABSTRACT

Learning health systems use routinely collected electronic health data (EHD) to advance knowledge and support continuous learning. Even without randomization, observational studies can play a central role as the nation's health care system embraces comparative effectiveness research and patient-centered outcomes research. However, neither the breadth, timeliness, volume of the available information, nor sophisticated analytics, allow analysts to confidently infer causal relationships from observational data. However, depending on the research question, careful study design and appropriate analytical methods can improve the utility of EHD.

The introduction to a series of four papers, this review begins with a discussion of the kind of research questions that EHD can help address, noting how different evidence and assumptions are needed for each. We argue that when the question involves describing the current (and likely future) state of affairs, causal inference is not relevant, so randomized clinical trials (RCTs) are not necessary. When the question is whether an intervention improves outcomes of interest, causal inference is critical, but appropriately designed and analyzed observational studies can yield valid results that better balance internal and external validity than typical RCTs. When the question is one of translation and spread of innovations, a different set of questions comes into play: How and why does the intervention work? How can a model be amended or adapted to work in new settings? In these “delivery system science” settings, causal inference is not the main issue, so a range of quantitative, qualitative, and mixed research designs are needed.

We then describe why RCTs are regarded as the gold standard for assessing cause and effect, how alternative approaches relying on observational data can be used to the same end, and how observational studies of EHD can be effective complements to RCTs. We also describe how RCTs can be a model for designing rigorous observational studies, building an evidence base through iterative studies that build upon each other (i.e., confirmation across multiple investigations).

ⁱGeorgetown University, ⁱⁱUniversity of Minnesota, ⁱⁱⁱJohns Hopkins University, ^{iv}Kaiser Permanente Center for Health Research, ^vBaylor Health Care System, ^{vi}Mathematica Policy Research

Introduction

A learning health system is one in which “science, informatics, incentives, and culture are aligned for continuous improvement and innovation, with best practices seamlessly embedded in the care process, patients and families active participants in all elements, and new knowledge captured as an integral by-product of the care experience”.¹ According to Fineberg,² these systems are continuously committed to improvement: they are engaged in discovery, innovation, and implementation, embedding research into the process of care. In particular, learning health systems, which can range from national systems such as Medicare to integrated health care delivery systems to physician practice networks such as practice-based research networks to individual hospitals, use routinely collected electronic health data (EHD) to advance knowledge and support continuous learning. Indeed, the increasing popularity of the term “learning health care system” signals the broad acceptance of the idea that routinely collected clinical data can—and indeed should—be used to advance knowledge and support continuous learning so as to improve patient care.³

Okun and colleagues,⁴ for instance, demonstrate how EHD can help to improve disease monitoring and tracking; better target medical services for improved health outcomes and cost savings; help inform both patients and clinicians to improve how they make decisions during clinical visits; avoid harm to patients and unnecessary costs associated with repeat testing and delivery of unsuccessful treatments; and accelerate and improve the use of research in routine medical care to answer medical questions more effectively and efficiently.

“Big data” – which includes but is not limited to EHD – has many uses in a learning health system.^{5,6,7,8} Some big data enthusiasts even seem to suggest that with enough data RCTs, long viewed as the gold

standard in health research, are no longer necessary.⁹ Schneeweiss¹⁰ argues that most of this can be automated, without losing validity, but Dahabreh and Kent¹¹ agree with the need for this kind of analysis but are less sanguine about the potential.

However, despite the new tools available for EHD, none of this is simple. Whether health system data are used to manage the care of individuals, to carry out quality improvement (QI) studies, or to conduct comparative effectiveness research (CER), data in individuals’ medical records must be accurate and complete; and the collection of records available for analysis must be reasonably representative of the population served. Without randomizing a large number of subjects to a treatment or intervention of interest, research results can suffer bias, especially by confounding by factors not recorded in existing electronic health records (EHR). Indeed, some seem to feel that without RCTs, we know nothing with certainty.¹² However, depending on the research question, careful study design and appropriate analytical methods can improve the utility of EHD in a learning health system and even yield convincing causal inferences.

Project goals

The goals of this set of papers (See Box 1) are to (1) illustrate how existing EHD data can be used to improve performance in learning health systems, (2) describe how to frame research questions to use EHD most effectively, and (3) determine the basic elements of study design and analytical methods that can help to ensure rigorous results in this setting. We explore how researchers can balance the rigor and internal validity of RCTs with the relevance and external validity of real world observational studies based on EHD. There are many existing publications addressing one or more of these topics, and most health professionals are aware of some, but not all, of these methods and approaches.



Box 1. Series on Analytic Methods to Improve the Use of Electronic Health Data in a Learning Health System

This is one of four papers in a series of papers intended to (1) illustrate how existing electronic health data (EHD) data can be used to improve performance in learning health systems, (2) describe how to frame research questions to use EHD most effectively, and (3) determine the basic elements of study design and analytical methods that can help to ensure rigorous results in this setting.

- Paper 1, this paper, focuses on clarifying the research question, including whether assessment of a causal relationship is necessary; why the randomized clinical trial (RCT) is regarded as the gold standard for assessing causal relationships, and how these conditions can be addressed in observational studies.
- Paper 2, “Design of observational studies,”¹³ addresses how study design approaches, including choosing appropriate data sources, methods for design and analysis of natural and quasi-experiments, and the use of logic models, can be used to reduce threats to validity in assessing whether interventions improve outcomes of interest.
- Paper 3, “Analysis of observational studies,”¹⁴ describe how analytical methods for individual-level electronic health data EHD, including regression approaches, interrupted time series (ITS) analyses, instrumental variables, and propensity score methods, can be used to better assess whether interventions improve outcomes of interest.
- Paper 4, “Delivery system science,”¹⁵ addresses translation and spread of innovations, where a different set of questions comes into play: How and why does the intervention work? How can a model be amended or adapted transported to work in new settings? In these settings, causal inference is not the main issue, so a range of quantitative, qualitative, and mixed research designs are needed.

This paper will not attempt to serve as a textbook or describe these approaches in detail. Rather, it presents these methods in a consistent framework rather than provide detailed information on each topic. The intended audience is health system researchers, analysts, and managers who want to understand the range of methods that have been developed for the analysis of EHD, their strengths and weaknesses, and the circumstances in which it is appropriate to apply them. Even though the examples are drawn mainly from large-scale research and evaluation studies where these methods have mostly been employed, the principles apply to topics relevant to smaller health systems and individual hospitals as well.

Organization of the series of papers

This set of papers begins with a discussion of the kind of research questions that EHD can help address, noting how different kinds of evidence and assumptions are needed for each. We argue that when the question involves describing the current (and likely future) state of affairs, causal inference – and hence RCTs – is not relevant. When the question is whether an intervention improves outcomes of interest, causal inference is critical, but appropriately designed and analyzed observational studies can yield valid results that better balance internal and external validity than typical RCTs. When the question is one of translation and spread

of innovations, a different set of questions comes into play: How and why does the intervention work? How can a model be amended or adapted to work in new settings? In these “delivery system science” settings, causal inference is not the main issue, so a range of quantitative, qualitative, and mixed research designs are needed. We then describe why RCTs are regarded as the gold standard for assessing cause and effect, how alternative approaches relying on observational data can be used to the same end, and how observational studies of EHD can be effective complements to RCTs. We also describe how RCTs can be a model for designing rigorous observational studies, building an evidence base through iterative studies that build upon each other (i.e., confirmation across multiple investigations).

We recognize that randomized studies are and will continue to be critical for learning health systems. This includes studies with randomization conducted at the individual or group level, using stepped-wedge¹⁶ and delayed-start¹⁷ designs, as well as pragmatic clinical trials¹⁸ and registry-based RCTs.¹⁹ Although EHD increasingly is being used in such studies, and even though we draw lessons and inspiration from these designs, randomized studies *per se* are beyond the scope of this paper. We also do not address measurement or data quality issues.

The current paper is intended as an introduction to the series, focusing on framing the research question, describing why RCTs are the gold standard for establishing cause and effect, and exploring how elements of RCTs can be used to strengthen causal inference in observational studies. Additionally, to further assess the rigor and appropriateness of observational studies, we point researchers to the literature on within-study comparisons. This involves the comparison of quasi-experimental designs to RCTs to assess whether, and in under what circumstances, observational studies and RCT have similar results. The other three papers

in the series explore different perspective on the use of observational data. The second paper²⁰ describes study design and analysis approaches for observational data that can be used to evaluate interventions, drawing primarily on epidemiology, social sciences, and the program evaluation literature. This includes logic models and methods for design and analysis of natural and quasi-experiments. The third paper,²¹ drawing primarily on statistics and econometrics, discusses methods for analysis of individual-level data including regression approaches, interrupted time series (ITS) analyses, instrumental variables, and propensity score methods. The final paper²² addresses methods from delivery system science for the evaluation of health care improvement initiatives. As indicated above, the questions are different, but the design and analytic approaches discussed in the second and third papers, along with other more qualitative and mixed methods such as realist and other theory-based evaluation approaches, can be useful.

What are the research questions?

Before we can even consider the most appropriate analytical methods for design and analysis of observational studies relying on EHD, we must begin by clarifying the questions to be addressed. Observational EHD can be used for a number of purposes. When it is used in place of an RCT to draw inferences about an intervention’s effects, great caution is needed in the selection of a counterfactual, the econometric approach used in estimation, and the interpretation.

For other questions, assessing causality is less central, and the use of observational data less contentious. For instance, one common type of question learning health systems address relates to the current (and likely future) situation for patients, providers, or health care systems. For instance, the question might be about:



- The magnitude of a disease or condition; for instance, the prevalence of diagnosed and undiagnosed diabetes among adults in medical practice X or in county Y.
- The occurrence, timing, and patterns of adverse events; for instance, without changing practice, how many new cases of hospital acquired infections should we expect next year? Or, how does the rate of central line-associated blood stream infections (CLASBI) compare to peer institutions?
- Cost and utilization of health care; for example, who are the most intensive users of health care (“super utilizers”) in the system?

For such descriptive questions, causal inference is simply not relevant. Indeed, because they relate to actual practice, existing EHD will likely provide more appropriate answers than RCTs, which are typically designed to yield precise answers in ideal contexts. Because causal inference is not an issue, these questions are beyond the scope of this series of papers.

The second type of question essentially asks whether an intervention “works,” that is, improves outcomes of interest. For example, since the 2006 Massachusetts health care reform was a model for the Affordable Care Act enacted four years later, it is important to ask about the Massachusetts reform’s impact on health care utilization, morbidity, and mortality. Learning health systems are also focused on safety, asking questions such as, what is the risk of intussusception after vaccination with second-generation rotavirus vaccines? Questions about whether an intervention “works” must also ask “compared to what?” For instance, is an intervention compared to usual care or a placebo or is it a question of comparative effectiveness, such as “what is the risk of short-term mortality associated with initiation of conventional vs. atypical

antipsychotic medication (APM)?” It is important to distinguish between effectiveness and efficacy, for instance, “what is the relative risk of mortality due to initiation of a conventional vs. atypical APM in actual practice?” (As compared to in some idealized setting that might, for example, just include large academic medical centers). For this type of question causal inference is critical, but appropriately designed and analyzed observational studies can yield valid results that better balance internal and external validity than RCT.²³ Indeed for issues of comparative effectiveness as opposed to efficacy, observational studies of EHD offer many advantages over simply RCTs.²⁴ Study designs and analytical methods for these questions are the major focus of this paper.

A third type of question, common in health care improvement efforts, arises when the focus is on the translation and spread of interventions from the institutions where they were developed to others throughout the health care system. For example, consider a patient-centered medical home (PCMH) approach to managing patients with diabetes. These complex interventions might rely on combinations of team-based care, health information technology and registry functionality, care coordination and management, and quality-adjusted financial incentives, with the specific versions of these components depending on the context in which the intervention is employed. For instance, Psek and colleagues²⁵ describe the implementation of learning health system principles in the Geisinger Health System, identifying “evaluation and methodology” (activities and methodological approaches needed to identify, implement, measure, and disseminate learning initiatives) as one of nine learning health system framework components. In a similar description of Kaiser Permanente’s approach to a learning health system, Schilling and colleagues²⁶ identify “real-time sharing of meaningful performance data” as one of six building blocks.

In this “delivery system science” setting, questions about whether or not interventions could improve outcomes are mostly settled. Accordingly, in these settings attention is devoted to not so much whether the intervention works, but rather (1) how and why the intervention works, (2) what works for whom and in what contexts, and (3) how a model can be amended to work in new settings.²⁷ Because the questions are different, the design and analysis approaches discussed in the second and third papers of this series, along with other more qualitative and mixed methods approaches such as realist and other theory-based evaluation methods are necessary. The main point is to appropriately match the design to the research question. Constraints related to context, capacity, and operational demands may require that we not consider certain design options—such as the case when randomization is not feasible or is ethically inappropriate. Further, operational considerations may require trade-offs in implementation strategies, moving from a randomized cluster trial design to a phased, time-delayed approach. Learning health systems engaging in delivery science must strive for rigor and researchers must balance the science with necessary tradeoffs necessary to care delivery operations in the pursuit of generating operationally relevant evidence.

RCTs as a model for observational studies using EHD

RCTs are critical in health research because randomization with large numbers of subjects is the best method (i.e., it requires the fewest and most defensible assumptions) to ensure that an association between a treatment and a health outcome represents a cause and effect relationship. This is because, when carried out well, randomization ensures that the groups receiving different treatments differ at baseline *only* on those treatments. So, in a “clean” RCT with few

or no complications, if differences in outcomes are observed those effects must be due to the treatments being compared, and not to any pre-existing differences between the groups. This quality is what makes randomization the “gold standard” for questions of efficacy, or in terms of the questions above, for determining whether an intervention “works.”

However, aside from pragmatic trials, most RCTs are limited by restrictions on subjects to obtain homogeneity (e.g., excluding patients with comorbidities) or due to their design and where they are carried out (e.g., in large medical centers rather than community hospitals), and as a result their participants often do not represent real-world patient populations. An additional drawback is that RCTs are often quite expensive. Because of their expense, RCTs often have small sample sizes, which limits their ability to detect rare adverse effects and study heterogeneity of treatment effects. RCTs also typically have short durations, limiting ability to discern long-term consequences. Thus, RCTs are generally good for internal validity (estimating effects for the research subjects in the study), but weak for external validity (estimating effects for some target population of interest). RCTs also have other limitations in terms of size and cost, and face complexities such as non-compliance or missing data, which make alternative designs sometimes attractive.

Observational studies using existing EHD offer the opportunity to investigate interventions and outcomes—often at lower costs and larger scale. In particular, Fleurence and colleagues²⁸ write that observational studies can play a central role as the nation’s health care system embraces comparative effectiveness research (CER). In fact, the field is moving towards consensus that future CER evidence is as likely to be drawn from the analysis of large observational databases as it is from clinical



trials.^{29,30,31,32,33} ARRA investments addressed the need for improved data infrastructure for CER with its investment in administrative and clinical datasets,³⁴ much of which is suited for observational studies. Given these investments, additional efforts are needed to educate researchers not only about the availability of new data sources, but also about how best to apply methods to match their research questions and data.³⁵ In their review of ARRA-funded data infrastructure projects, O'Day et al.³⁶ found that about half of all data infrastructure projects sought to leverage EHR data sources for CER.

More recently, other initiatives have picked up where these efforts have left off. For instance, PCORnet, the National Patient-Centered Clinical Research Network, is an initiative of the Patient-Centered Outcomes Research Institute (PCORI). It is designed to make it faster, easier, and less costly to conduct clinical research than is now possible by harnessing the power of large amounts of health data and patient partnerships.³⁷ One of the first PCORnet observational studies is examining the relationship between antibiotic use and weight gain during childhood.³⁸ Building on the work of the Observational Medical Outcomes Partnership (OMOP), the Observational Health Data Sciences and Informatics (OHDSI) is a collaborative that strives to bring out the value of observational health data through large-scale analytics. Its research community enables active engagement across multiple disciplines (e.g., clinical medicine, biostatistics, computer science, epidemiology, life sciences) and spans multiple stakeholder groups (e.g., researchers, patients, providers, payers, product manufacturers, regulators).^{39,40} The Food and Drug Administration's (FDA) Sentinel Initiative enhances the FDA's ability to proactively monitor the safety of medical products after they have reached the market by rapidly and securely accessing information from large amounts of electronic health

care data, such as electronic health records (EHR), insurance claims data and registries, from a diverse group of data partners.⁴¹ OptumLabs, a collaborative health care research and innovation center and its private- and public-sector partners, also undertakes initiatives that harness the power of existing electronic health data to improve patient care and outcomes.⁴² Existing EHD is also essential for the Center for Medicare and Medicaid Innovation's new, rapid-cycle approach to evaluation, which aims to deliver frequent feedback to providers in support of continuous QI, while rigorously evaluating the outcomes of each model tested.⁴³

Strengths and weaknesses of observational studies

The strengths of observational studies that can be conducted using EHD are the large, diverse populations "under observation" and the relatively (and increasingly) complete information in electronic medical records (EMR) and other EHD on treatments administered and health outcomes experienced. In addition to representing what happens in the "real world," as needed to estimate effectiveness, observational data representing large populations are increasingly available, making it possible to estimate how treatment effects vary across subgroups defined by demographic factors, geography, disease severity and comorbidity, physiological variables, and patient reported measures. The data are already in electronic form, permitting more rapid analyses and possibly lower costs compared to typical RCTs. In addition, reporting bias may be minimized because the data are collected for operational rather than research purposes. According to the Network for Excellence in Health Innovation,⁴⁴ evidence from "real world" practice and utilization – outside of clinical trials – is seen as a way to tailor health care decision making more closely to the characteristics of individual patients. While RWE will not supplant the randomized clinical trial (RCT), appropriate adoption

of RWE by sponsors of new drugs and devices and regulators could streamline or supplement data from RCTs. Analysis of real world data can help expedite generation of research hypotheses that sharpen the focus of clinical research and may augment conventional RCT data with data from patients whose diversity reflects real world practice. Analysis of long-term observation of patient outcomes from the use of innovations in real world settings generates further insight on safety and efficacy.

The major weakness of observational studies is the lack of randomization, creating the possibility, or indeed likelihood, of some degree of confounding and/or selection bias. EHD systems designed for health care operations typically lack complete information on the confounders (e.g. the patients' home situation and caregiver status) needed to adjust for treatment and control group differences. In addition, the data available are typically not of "research grade," i.e. not collected with the completeness and accuracy typical of a carefully planned RCT. There are also ethical considerations – primarily privacy and confidentiality – that limit researchers' access to existing data for health research purposes⁴⁵ when researchers are not employed by the health care system.

Using principles of RCTs to design better observational studies

Learning health systems often require evidence of a cause and effect relationship; simply observing an improvement in targeted outcomes, for example, does not tell us whether a given patient improved because of a particular treatment, or whether they would have improved even without such intervention. Observational studies are generally regarded as establishing association but not necessarily causation. But in certain circumstances, well-designed observational studies can indeed contribute to, if not provide, evidence of a cause

and effect relationship. The key issue is the assumptions being made, and assessing their validity in each study. Just as in RCTs that face complexities such as non-compliance or missing data, inference from observational designs requires data and assumptions. In part to help motivate thinking regarding the careful design and analysis of observational studies we consider the two major reasons why RCTs with a large number of subjects are considered the gold standard in health research.

First, RCTs are particularly useful in studying the efficacy of medications, devices, and health services provided to individuals – and have earned their reputation – because individual patient outcomes vary and are unpredictable. Only a fraction respond favorably to even the most efficacious medications, and some people get better even without treatment. Moreover, confounding bias is often difficult to avoid in observational studies; income, education, insurance, and many other factors are often common causes of both treatments and health outcomes. The key advantage of an RCT is that randomization balances all potential confounders – measured and unmeasured – between comparison groups. RCTs thus provide good information about what would have happened to individuals had they instead gotten the other treatment condition (known as a "counterfactual"). Counterfactuals are unobservable outcomes that represent what would have happened if a treatment or exposure had not been applied. Through randomization of large numbers of subjects, RCTs produce observable data to substitute for the unobservable counterfactual conditions. Due to randomization in RCTs, the treatment and control groups are equivalent at baseline on both observable and unobservable characteristics, the treatment group outcomes are a good indication of what would have happened to the control group had the control group members instead gotten the treatment of interest, thus allowing a



reliable estimation of causal effects. As we discuss in the second and third papers in this series, both retrospective and prospective quasi-experimental designs provide matched and unmatched control options without randomization.⁴⁶

Sometimes it is not clear how the desired but unobservable counterfactual might be found in the real world. For example, under what conditions could we observe “black” persons being “white?” Such phenomena complicate the estimation of so-called race effects. In other studies, however, the counterfactual is clear. With tongue in cheek, for instance, Smith and Pell⁴⁷ challenge the conventional wisdom that parachutes can “prevent death and major injury after gravitational challenge,” i.e., jumping out of an airplane, because of the lack of RCTs on the subject. This is humorous because everyone knows what the counterfactual to the use of a parachute would be. But unlike the use of medications and other individual-level interventions, the counterfactual for some systems-level changes can be quite clear. Consider the Keystone Initiative study by a collection of Michigan intensive care units (ICUs) of the impact of a simple checklist on health care acquired infections. When the authors found that the median rate of infections at a typical ICU dropped from 2.7 per 1,000 patients to zero after three months, which was sustained for 15 months of follow-up,⁴⁸ no one thought that the rate would have dropped so precipitously without the intervention.

Second, RCTs help establish a cause and effect relationship because they address the three criteria for a “contributory cause”.^{49,50} First, the cause must precede the effect. This is true by definition in an RCT because the outcomes are observed after the subjects are randomized. Second, there must be an individual-level association between cause and effect, or that there is a difference in outcomes between those who receive the treatment and those who do not. Randomization ensures that there are

no systematic differences between treatment and control groups at the time of the randomization/treatment assignment, even in unobserved variables. Finally, altering the cause results in a change in the effect. Susser describes this third criterion as “direction,” which is indicated by the presence of consequential change. An active agent (such as the use of a medication) itself changes and, in turn, is shown to change the outcome. With a static determinant, the effect changes in consequence of a change or shift in a prior condition, such as having diabetes as a prior condition.⁵¹ RCTs feature a deliberate allocation of subjects to treatment and control groups. In RCTs this allocation is done on a random basis but it is worth noting that deliberate allocation is also a feature of natural or quasi experiments as well as in health care QI efforts. This allocation, however, is neither under the control of the researcher or done at random.

A third strength of RCTs is that they are carefully planned. Researchers should and do use lessons about RCTs to help design observational studies with just as much care. William Cochran, author of the seminal book, *Planning and Analysis of Observational Studies*,⁵² suggested that researchers carefully design observational studies by asking how the study would be conducted if it were possible to do it by controlled experimentation,⁵³ and recent authors^{54,55} have reinforced those ideas. In particular, observational studies should aim to replicate the following key features of an experiment:

- Clear definition of treatment and comparison conditions,
- Clear inclusion and exclusion criteria for the study,
- Methods to adjust for differences in observed characteristics between groups as a way to mitigate the inherent differences, and
- Identifying when evidence of an association might suggest a causal relationship.

Assessing causality in observational studies

The statistical methods for analysis of individual-level data discussed in the third paper in this series⁵⁶ all relate to Cochran's third point. Regarding the last point, as early as the 1950s epidemiologists have used the "Bradford Hill criteria" to determine whether existing observational data – taken together with an understanding of the related science – are sufficient to say that there is a cause and effect relationship. These criteria developed by the British statistician Sir Austin Bradford Hill are summarized as follows:

1. Strength of association between risk factor (cause) and disease (health effect) measured by the risk ratio (RR), odds ratio (OR), or other measure of association,
2. Consistency of association (studies in different settings produce similar results),
3. Dose-response relationship (the risk of disease increases with level or extent of exposure)
4. Temporal relationship (the cause precedes the effect),
5. Biological plausibility (based on scientific theory; timing and magnitude compatible with effect),
6. Coherence (consistency with pre-existing theory and other knowledge), and
7. Specificity (the cause is not associated with other health effects).

Hill did not intend that these criteria be used in an algorithmic way, for instance that all need to be met before an association is judged to be causal, but rather that they serve as reminders of the kinds of evidence needed to make that judgment.

For example, consider the evidence regarding smoking and lung cancer at the time of the first U.S. Surgeon General's report issued in 1964 (Cochran served as the lead statistician). No one today questions the causal relationship between smoking and lung cancer, even though this has never been

tested in an RCT in humans. The evidence available in 1964 showed that the strength of association was high, with estimated RR of greater than 10 in a range of studies. This association was also consistent; elevated RRs were found in case-control and cohort studies, different populations, and so on. There was also evidence of a dose-response relationship in epidemiological studies showing a higher risk of lung cancer with increasing time smoked and number of cigarettes. The temporal relationship was clear – the epidemiological studies compared current cancer to earlier smoking – and coherence demonstrated by increased risk at sites other than the lung.

The biological plausibility of the relationship was demonstrated through randomized experimental studies in animals, higher risks at anatomical sites more exposed to tobacco smoke, and what was known at the time about the causes of cancer. The specificity of the relationship is the only criteria not met. In this example the evidence is strong, and the public health need for action was compelling. In most other cases faced by learning health systems, the call will be much harder.

In addition to carefully designing observational studies, the within-study comparison literature can provide much guidance about observational study methods, as it is generating empirical evidence of the degree of correspondence in estimated treatment effects between observational studies and RCTs. This literature usually involves a head-to-head comparison of effects estimated in an RCT to that of an observational study, although the focus has recently started moving to studying the circumstances under which the replication is most likely, in addition to assessing correspondence between the two. Most of the past within-study comparisons dealt with interventions in job training and education with relatively few studies in health,⁵⁷ but the literature is rapidly expanding. Much remains to be learned, especially about the observational



study designs that are much more common place because they are more feasible, but also more problematic in terms of the assumptions made, such as propensity score methods.

References

1. Institute of Medicine, Best care at lower cost. Washington (DC): National Academy Press. 2012.
2. Fineberg H. Research and health policy: where the twain shall meet. Presentation at: AcademyHealth Annual Research Meeting; 2014 Jun 8-10; San Diego, CA.
3. Longhurst CA, Harrington RA, Shah NH. A 'green button' for using aggregate patient data the point of care. *Health Affairs*. 2014 Jul; 33(7):1229-1235.
4. Okun S, McGraw D, Stang P, Larson E, Goldmann D, Kupersmith J, Filart R, Robertson RM, Grossmann C, Murray M. Making the case for continuous learning from routinely collected data [Internet]. Washington (DC): Institute of Medicine; 2013 Apr 15 [cited 2014 Dec 17]. Available from: <http://www.iom.edu/Global/Perspectives/2013/MakingtheCaseforContinuousLearning.aspx>
5. Brennan N, Oelschlaeger A, Cox C, Tavenner M. Leveraging the big-data revolution: CMS is expanding capabilities to spur health system transformation. *Health Aff*. 2014, 33(7):1195-1202.
6. Krumholz HM. Big data and new knowledge in medicine: the thinking, training, and tools needed for a learning health system. *Health Aff*. 2014, 33(7):1163-1170.
7. Roski J, Bo-Linn GW, Andrews TA. Creating value in health care through big data: opportunities and policy implications *Health Aff*. 2014, 33(7):1115-1122.
8. Wallace PJ, Shah ND, Dennen T, Bleicher PD, Crown WH. Optum Labs: building a novel node in the learning health care system. *Health Aff*. 2014, 33(7):1187-1194.
9. Mearian L. How big data will save your life [Internet]. Framingham (MA): Computerworld; 2013 Apr 25 [cited 2014 Dec 17]. Available from: http://www.computerworld.com/s/article/9238593/How_big_data_will_save_your_life
10. Schneeweiss S. Learning from big health care data. *N Engl J Med*. 2014 Jun 5; 370(23): 2161-2163.
11. Dahabreh I, Kent DM. Can the learning health care system be educated with observational data? *JAMA*. 2014 Jul; 312(2): 129-130.
12. Kolata G. Method of study is criticized in group's health policy tests [Internet]. *New York Times*. 2014 Feb 3:Section. A:1.
13. Stoto MA, Oakes M, Stuart EA, Priest E, Savitz L. Analytical methods for a learning health system: 2. Design of observational studies. *eGEMs (Generating Evidence & Methods to improve patient outcomes)*. 2016;5(1):29.
14. Stoto MA, Oakes M, Stuart EA, Brown R, Zurovac J, Priest E. Analytical methods for a learning health system: 3. Analysis of observational studies. *eGEMs (Generating Evidence & Methods to improve patient outcomes)*. 2016;5(1):30.
15. Stoto MA, Parry G, Savitz L. Analytical methods for a learning health system: 4. Delivery system science. *eGEMs (Generating Evidence & Methods to improve patient outcomes)*. 2016;5(1):31.
16. Fan E, Laupacis A, Pronovost PJ, Guyatt GH, Needham DM. How to use an article about quality improvement. *JAMA*. 2010 Nov 24; 304(20): 2279-2287.
17. Velengtas, P, Mohr P, Messner DA, editors. Making informed decisions: assessing the strengths and weaknesses of study designs and analytic methods for comparative effectiveness research. Washington (DC): National Pharmaceutical Council; 2014. 36 p.
18. Thorpe KE, Zwarenstein M, Oxman AD, Treweek S, Furberg CD, Altman DG, Tunis S, Bergel E, Harvey I, Magid DJ, Chalkidou K. A pragmatic explanatory continuum indicator summary (PRECIS): a tool to help trial designers. *CMAJ*. 2009 May 12; 180(10): E47-E57.
19. Lauer MS, D'Agostino RB Sr. The randomized registry trial—the next disruptive technology in clinical research? *N Engl J Med*. 2013 Oct 24; 369(17): 1579-1581.
20. Stoto MA, Oakes M, Stuart EA, Priest E, Savitz L. Analytical methods for a learning health system: 2. Design of observational studies. *eGEMs (Generating Evidence & Methods to improve patient outcomes)*. 2016;5(1):29.
21. Stoto MA, Oakes M, Stuart EA, Brown R, Zurovac J, Priest E for the EDM Forum Methods Collaborative. Analytical methods for a learning health system: 3. Analysis of observational studies. *eGEMs (Wash DC)*. Forthcoming 2017.
22. Stoto MA, Parry G, Savitz L. Analytical methods for a learning health system: 4. Delivery system science. *eGEMs (Generating Evidence & Methods to improve patient outcomes)*. 2016;5(1):31.
23. Imai K, King G, Stuart EA. Misunderstandings between experimentalists and observationalists about causal inference. *J. R. Statist. Soc. A*. 2008; 171(2): 481-502.
24. Fleurence RL, Naci H, Jansen JP. The critical role of observational evidence in comparative effectiveness research. *Health Aff (Millwood)*. 2010 Oct; 29(10): 1826-1833.
25. Psek WA, Stametz RA, Bailey-Davis LD, Davis D, Jonathan D, Faucett WA, Henninger DL, Sellers DC, Gerrity G. Operationalizing the Learning Health Care System in an Integrated Delivery System. *eGEMs (Generating Evidence & Methods to improve patient outcomes)* 2015, 3(1), Article 6. Available at: <https://egems.academyhealth.org/articles/abstract/10.13063/2327-9214.1122/>.
26. Schilling L, Dearing JW, Staley P, Harvey P, Fahey L, Kuruppu F. Kaiser Permanente's performance improvement system, part 4: creating a learning organization. *The Joint Commission Journal on Quality and Patient Safety* 2011, 36(12):
27. Parry GJ, Carson-Stevens A, Luff DF, McPherson ME, Goldmann DA. Recommendations for evaluation of health care improvement initiatives. *Acad Pediatr*. 2013 Nov-Dec; 13(6 Suppl): S23-S30.
28. Fleurence RL, Naci H, Jansen JP. The critical role of observational evidence in comparative effectiveness research. *Health Aff*. 2010 Oct; 29(10): 1826-1833.

29. U.S. Department of Health and Human Services, Federal Coordinating Council for Comparative Effectiveness Research. Report to the President and the Congress [Internet]. Washington: U.S. Department for Health and Human Services; 2009 Jun 30 [cited 2014 Dec 17]. 73 p. Available from: http://www.med.upenn.edu/sleepctr/documents/FederalCoordinatingCouncilforCER_2009.pdf
30. Grossman C, Alper J, editors. Observational studies in a learning health system: workshop summary; 2013 Apr 25-26; Washington, DC. Washington: National Academies Press; 2013. 134 p.
31. Dreyer NA, Tunis SR, Berger M, Ollendorf D, Mattox P, Gliklich R. Why observational studies should be among the tools used in comparative effectiveness research. *Health Aff.* 2010 Oct; 29(10): 1818-1825.
32. Armstrong K. Methods in comparative effectiveness research. *J Clin Oncol.* 2012 Dec 1; 30(34): 4208-4214.
33. Ryan P. Statistical challenges in systematic evidence generation through analysis of observational healthcare data networks. *Stat Methods Med Res.* 2013 Feb; 22(1): 3-6.
34. O'Day B, Kieffer T, Forrestal S, Esposito D. American Recovery and Reinvestment Act investments in data infrastructure. *J Comp Eff Res.* 2014 Nov; 3(6): 591-600.
35. Zurovac J, Esposito D. Lessons from comparative effectiveness research methods development projects funded under the Recovery Act. *J Comp Eff Res.* 2014 Nov; 3(6): 601-607.
36. O'Day B, Kieffer T, Forrestal S, Esposito D. American Recovery and Reinvestment Act investments in data infrastructure. *J Comp Eff Res.* 2014 Nov; 3(6): 591-600.
37. PCORnet, the National Patient-Centered Clinical Research Network [Internet]. Washington (DC): Patient Centered Outcomes Research Institute. [cited 2016 Sep 3] Available from: <http://pcor.net/org/>
38. PCORnet Obesity Observational Study: Short- and Long-term Effects of Antibiotics on Childhood Growth. [Internet]. Washington (DC): Patient Centered Outcomes Research Institute. [cited 2016 Sep 3] Available from: <http://www.pcori.org/research-results/2015/pcor-net-obesity-observational-study-short-and-long-term-effects-antibiotics>
39. Observational Health Data Sciences and Informatics. [Internet]. Observational Health Data Sciences and Informatics. [cited 2016 Sep 3] Available from: <http://www.ohdsi.org/>
40. Hripcsak G, Ryan PK, Duke JD, Shah NH, Park RW, Huserc V, Suchard MA, Schuemie MJ, DeFalco FJ, Perotte A, Banda JM, Reich CG, Schilling LM, Matheny ME, Meeker D, Pratt N, Madigan D. Characterizing treatment pathways at scale using the OHDSI network. *PNAS* 2016 Jul 5; 213(27):7329-7336.
41. FDA's Sentinel Initiative. [Internet]. Observational Health Data Sciences and Informatics. Silver Spring (MD): U.S. Food and Drug Administration. [cited 2016 Sep 3] Available from: <http://www.fda.gov/Safety/FDAsSentinelInitiative/default.htm>.
42. Wallace PJ, Shah ND, Dennen T, Bleicher PD, Crown WH. Optum Labs: building a novel node in the learning health care system. *Health Aff.* 2014, 33(7):1187-1194.
43. Shrank W. The Center for Medicare and Medicaid Innovation's blueprint for rapid-cycle evaluation of new care and payment models. *Health Aff.* 2013 Apr; 32(4): 807-812.
44. The Network for Excellence in Health Innovation. Real World Evidence: A New Era for Health Care Innovation [Internet]. Cambridge (MA): The Network for Excellence in Health Innovation; 2015 [cited 2016 Oct 3]. Available from: http://www.nehi.net/writable/publication_files/file/rwe_issue_brief_final.pdf
45. Gostin LO, Nass S. Reforming the HIPAA Privacy Rule: safeguarding privacy and promoting research. *JAMA.* 2009 Apr 1; 301(13): 1373-1375.
46. Campbell DT, Stanley JC. Experimental and quasi-experimental designs for research. Chicago: Rand McNally; 1963. 84 p.
47. Smith GCS, Pell JP. Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomized controlled trials. *BMJ.* 2003 Dec 20; 327(7429): 1459-1461.
48. Pronovost P, Needham D, Berenholtz S, Sinopoli D, Chu H, Cosgrove S, Sexton B, Hyzy R, Welsh R, Roth G, Bander J, Kepros J, Goeschel C. An intervention to decrease catheter-related bloodstream infections in the ICU. *N Engl J Med.* 2006 Dec 28; 355(26): 2725-2732.
49. Riegelman RK. Studying a study and testing a test: reading evidence-based health research. 6th ed. Baltimore: Lippincott Williams & Wilkins; 2003. 340 p.
50. Susser M. What is a cause and how do we know one? a grammar for pragmatic epidemiology. *American Journal of Epidemiology.* 1991, 133(7): 635-648.
51. Susser M. What is a cause and how do we know one? a grammar for pragmatic epidemiology. *American Journal of Epidemiology.* 1991, 133(7): 635-648.
52. Cochran WG. Planning and analysis of observational studies. New York: John Wiley & Sons; 1983. 145 p.
53. Cochran WG, Chambers SP. The planning of observational studies of human populations. *J. R. Statist. Soc. A.* 1965; 128(2): 234-266.
54. Rosenbaum PR. Choice as an alternative to control in observational studies. *Statistical Science.* 1999; 14(3): 259-304.
55. Hernan MA, Alonso A, Logan R, Grodstein F, Michels KB, Willett WC, Manson JE, Robins JM. Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. *Epidemiology.* 2008 Nov; 19(6): 766-779.
56. Stoto MA, Oakes M, Stuart EA, Brown R, Zurovac J, Priest E. Analytical methods for a learning health system: 3. Analysis of observational studies. *eGEMs (Generating Evidence & Methods to improve patient outcomes).* 2016;5(1):30.
57. Cook TD, Shadish WR, Wong VC. Three conditions under which experiments and observational studies produce comparable causal estimates: new findings from within-study comparisons. *Journal of Policy Analysis and Management.* 2008; 27(4): 724-750.