

12-30-2013

# Using Multifactorial Experiments For Comparative Effectiveness Research in Physician Practices with Electronic Health Records

Jelena Zurovac Ph.D, M.S.

*Mathematica Policy Research*, jzurovac@mathematica-mpr.com

Lorenzo Moreno Ph.D., M.Sc.

*Mathematica Policy Research*, lmoreno@mathematica-mpr.com

Jesse Crosson Ph.D.

*Mathematica Policy Research*, jcrosson@mathematica-mpr.com

Randall Brown Ph.D.

*Mathematica Policy Research*, rbrown@mathematica-mpr.com

*See next pages for additional authors*

Follow this and additional works at: <http://repository.edm-forum.org/egems>



Part of the [Health Information Technology Commons](#), and the [Policy Design, Analysis, and Evaluation Commons](#)

## Recommended Citation

Zurovac, Jelena Ph.D, M.S.; Moreno, Lorenzo Ph.D., M.Sc.; Crosson, Jesse Ph.D.; Brown, Randall Ph.D.; and Schmitz, Robert Ph.D. (2013) "Using Multifactorial Experiments For Comparative Effectiveness Research in Physician Practices with Electronic Health Records," *eGEMs (Generating Evidence & Methods to improve patient outcomes)*: Vol. 1: Iss. 3, Article 5.

DOI: <http://dx.doi.org/10.13063/2327-9214.1037>

Available at: <http://repository.edm-forum.org/egems/vol1/iss3/5>

This Methods Review is brought to you for free and open access by the the Publish at EDM Forum Community. It has been peer-reviewed and accepted for publication in eGEMs (Generating Evidence & Methods to improve patient outcomes).

The Electronic Data Methods (EDM) Forum is supported by the Agency for Healthcare Research and Quality (AHRQ), Grant 1U18HS022789-01. eGEMs publications do not reflect the official views of AHRQ or the United States Department of Health and Human Services.

---

# Using Multifactorial Experiments For Comparative Effectiveness Research in Physician Practices with Electronic Health Records

## Abstract

Two key challenges related to conducting comparative effectiveness research are the lack of available data and the lack of rigorous techniques for efficiently and quickly testing the effectiveness of the many possible ways of implementing components of care. The confluence of two things offers the promise of overcoming these challenges: (1) the increased adoption of electronic health records (EHRs), which can provide easier access to clinical information, and (2) burgeoning appreciation for an under-used but powerful statistical research and evaluation method for multifactor interventions known as multifactor experimental design.

The use of multifactorial experiments paired with EHR data has great potential to help providers conduct rapid-cycle comparative effectiveness research and examine alternative ways of implementing care. Its power is its ability to enable scientifically rigorous testing of many facets of care provision simultaneously in real-world settings where change is ongoing.

In this paper, we identify the opportunities for using efficient multifactorial designs and EHR data to evaluate quality-improvement efforts in physician practices. We illustrate the power of multifactorial designs through several examples relevant to physician practices with EHRs, such as evaluating clinical decision support features and studying components of a patient-centered medical home.

## Acknowledgements

We gratefully acknowledge the help of our colleague Arnold Chen in the preparation of this paper, as well as research assistance of Joe Brown.

## Keywords

Methods, comparative effectiveness, quality improvement, health information technology, data use and quality

## Disciplines

Health Information Technology | Policy Design, Analysis, and Evaluation

## Creative Commons License



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 License](https://creativecommons.org/licenses/by-nc-nd/3.0/).

## Authors

Jelena Zurovac, *Mathematica Policy Research*; Lorenzo Moreno, *Mathematica Policy Research*; Jesse Crosson, *Mathematica Policy Research*; Randall Brown, *Mathematica Policy Research*; Robert Schmitz, *Mathematica Policy Research*.

# Using Multifactorial Experiments for Comparative Effectiveness Research in Physician Practices with Electronic Health Record

\*Jelena Zurovac PhD, MS; Lorenzo Moreno PhD, MSc; Jesse Crosson PhD; Randall Brown PhD; Robert Schmitz PhD

## Abstract

Two key challenges related to conducting comparative effectiveness research are the lack of available data and the lack of rigorous techniques for efficiently and quickly testing the effectiveness of the many possible ways of implementing components of care. The confluence of two things offers the promise of overcoming these challenges: (1) the increased adoption of electronic health records (EHRs), which can provide easier access to clinical information, and (2) burgeoning appreciation for an under-used but powerful statistical research and evaluation method for multifactor interventions known as multifactor experimental design. The use of multifactorial experiments paired with EHR data has great potential to help providers conduct rapid-cycle comparative effectiveness research and examine alternative ways of implementing care. Its power is its ability to enable scientifically rigorous testing of many facets of care provision simultaneously in real-world settings where change is ongoing. In this paper, we identify the opportunities for using efficient multifactorial designs and EHR data to evaluate quality-improvement efforts in physician practices. We illustrate the power of multifactorial designs through several examples relevant to physician practices with EHRs, such as evaluating clinical decision support features and studying components of a patient-centered medical home.

## Introduction and Background

Two key challenges associated with conducting comparative effectiveness research in primary care practice redesign are the lack of readily available data and the lack of rigorous techniques for efficiently and quickly testing the effectiveness of the many possible ways of implementing components of care. The confluence of two things offers the promise of overcoming these challenges: (1) the increased adoption of electronic health records (EHRs), which can offer easier access to clinical information, and (2) burgeoning appreciation for an underused but powerful statistical research and evaluation method for multifactor interventions known as multifactor experimental design.

Multifactor experimental designs combine the rigor of experimental design with the ability to produce results on the effectiveness of alternate approaches to multicomponent interventions in a single experiment. As a result, the approach has great potential as a tool for fostering continuous improvement in health care quality and outcomes.

In multifactor experimental designs, researchers must specify in advance the alternate approaches to implementing key components that they wish to test. That contrasts with two-arm experimental designs, where only one component or one set of combined components is tested. Multifactor experimental design allows testing

under real-world conditions, enabling comparisons of routine and enhanced care without the need for a traditional control group, thus permitting rapid testing of quality improvement initiatives. With this approach, each provider is assigned to deliver to all of their patients a specific combination of enhanced or routine versions of the components of care being tested. Multifactor experimental design studies may make recruitment easier, as all experimental units are engaged in testing new enhancements: each unit is assigned to implement some enhanced care and some care alternatives. In this context, alternatives refer to different ways of implementing the components (e.g., different approaches to managing care transitions, or different protocols for screening for falls risk). We use the term “quality improvement” broadly, to denote any activities aimed at improving the quality of care provided in health care, regardless of the model of quality improvement used.

Multifactor experimental designs originated in agricultural experiments in England<sup>1</sup> and have been used extensively in manufacturing (in the automotive industry, for example) and in marketing (to study consumer preferences and choices). They have been used in some health care organizations for internal program-improvement assessments, but the only published studies in the health care arena are from the behavioral health field, and they are not widely known in other areas of health care research. The main goals of this

\*Mathematica Policy Research

paper are to introduce these powerful designs to the broader health research community and to illustrate their strength in rigorously analyzing multifactor interventions. An important auxiliary goal is to highlight the ways in which EHRs make it easier to conduct such studies and the ways that this design tool can be used to improve the meaningful use of EHRs.

To accomplish this, we focus on identifying the opportunities for using efficient designs and EHR data to assess quality improvement efforts in physician practices. More specifically, we focus on the use of efficient designs—those that offer the most value for research—because they require relatively few experimental units for a given number of tested intervention components. Examples of such designs include fractional factorial designs, which are based on factorial experiments that require a fraction of the experimental units to test the same number of components, and Plackett-Burman designs, which require very few experimental units to test many components<sup>2</sup>. Although the focus here is on efficient designs, it must be noted that inefficient designs have a place in research when relatively few interventions are analyzed in relation to the number of experimental units available.

We illustrate the usefulness of efficient multifactor experimental designs by offering two examples of how quality improvement could be implemented in practices that are using EHRs. We focus in this paper on the use of multifactor experimental designs for quality improvement in a set of physician practices in which each practice implements a combination of alternatives. This type of design can, however, also be used in a single care setting where individual clinicians are assigned to provide specific interventions.

## Using Multifactor Experimental Designs to Evaluate Quality Improvement in Physician Practices

Often initiated in individual practices, quality improvement is aimed at a wide variety of targets and multiple approaches to implementing particular facets of care. Heterogeneity inherent across and within health care settings must be embraced because it gives rise to various approaches to service provision, thereby enabling learning and facilitating quality improvement. However, to ensure the internal validity of a study, every well-designed experiment should study improvements in a well-defined setting or have sufficient statistical power to estimate whether the effectiveness of some interventions varies with characteristics of the setting.

A variety of organizations interested in advancing the quality of primary care—including those that provide medical care and those that set health care delivery policies—are likely to find efficient multifactor experimental design studies useful. For instance, large multispecialty practices or accountable care organizations could test the effect of various configurations of work-process redesign and technological interventions to improve quality of care in specific practice settings. Practice-based research networks provide another setting for testing quality improvements by leveraging their large numbers of affiliated

practices to examine various ways of implementing a specific quality improvement intervention. One of the most promising settings for conducting this type of research is multiple affiliated practices that all use the same EHR product, because that reduces the cost of data collection and could potentially facilitate quality improvement efforts that might require programming changes. However, in these three types of settings, heterogeneity at baseline must be examined to decide whether the evaluation should include only relatively homogeneous units or whether the units should be reorganized to reduce baseline heterogeneity.

Studies in primary care settings with EHRs could use efficient multifactor experimental designs to test the comparative effectiveness of multicomponent quality improvement interventions in patient care, including those delivered through existing EHR technologies. One area where efficient multifactorial designs and EHR systems show great promise for research and practice redesign is evaluating clinical decision support (CDS) alerts and reminders with the goals of improving the safety and quality of patient care and optimizing workflow. Another area in which efficient designs are especially promising are complex, multicomponent initiatives, such as those testing the patient-centered medical home (PCMH) model.<sup>3</sup> Results of efficient design studies in this area could address the need discussed by Peikes and colleagues to create the most effective—and cost-effective—“package” of medical home features.<sup>4</sup>

To harness the power of multifactor experimental design, one should leverage existing methods and approaches to quality improvement. For example, implementation science could be used to identify key components and alternatives before multifactor experimental design is employed to evaluate their effectiveness. During the implementation period, an implementation evaluation (1) assesses thoroughly and consistently the interventions that are implemented, and (2) documents difficulties encountered by implementers in operationalizing the interventions. This information can then be used to improve the interpretation of the impact evaluation findings, including operational information that may be useful to other providers in similar settings. The information can also provide a starting point for further quality improvement in those settings.

## Overview of Multifactor Experimental Design Features

We focus on multifactor experimental designs that test two alternatives for implementing the components of an intervention (referred to here as alternatives *a* and *b*). (There are more complex designs for testing three or more alternatives.) Examples of alternatives to test include different frequencies of displaying EHR alerts or reminders about evidence-based guidelines and different strategies to display them, such as to whom to display them. Various protocols that are followed during patient care can also be tested, such as means of communication with patients who have not received a given preventive service (e.g., telephone call versus email), or level of frequency (e.g., meeting with a patient once versus twice to explain key self-care issues).

The number of components that can be tested depends on the number of experimental units (such as the number of physician practices). Including all possible combinations in a full factorial design would require  $2^n$  practices, where  $n$  is the number of tested components (also referred to as factors). For example, 32 practices would be needed to implement each of the 32 possible combinations of five components. However, efficient designs require only a fraction of combinations of components to obtain unbiased estimates of the relative effects of alternatives  $a$  and  $b$  for each intervention component. (The relative effects of alternatives for a single component are referred to as “main effects”). The smaller the number of practices required to estimate a given number of main effects, the more efficient the design. For example, one such design prescribes a unique combination of alternatives that enables evaluators to estimate the main effects of five components with only eight practices. When relatively few components are being tested, less efficient designs (such as full factorial design) can help identify interactions between many or all components. For example, a full factorial design studying four components requires only 16 practices.

For any given number of practices participating in a study, published algorithms show various combinations of components to be implemented by the practices. The algorithms also illustrate possible tradeoffs between the number of components that can be tested and the extent to which the main effects are indistinguishable from effects of a combination of components (“interaction effects”). As a study tests more components relative to the number of practices, it becomes increasingly difficult to identify the main and interaction effects. Efficient designs enable estimation of main effects with few practices because they regard interactions of other components as negligible. The most efficient designs regard even two-way interactions as negligible, meaning that main effects are indistinguishable from the effects of two-way interactions. Because most components are likely to interact, the most efficient designs are best suited for screening components for further testing. To identify higher-order interactions, the study can include more practices and use a less efficient but more discriminating orthogonal design or reduce the number of tested components. See Zurovac and Brown<sup>5</sup> for further discussion. For example, in our study of care coordination strategies for patients served by special needs plans, the design required 24 care managers to study alternative ways to implement 11 approaches in care coordination; using this particular design, we were able to distinguish main effects from two-way interactions.

After selecting a particular design (algorithm), the evaluator randomly assigns (without replacement) one combination to each participating practice, then the practice administers the set of alternatives to all of its eligible patients. Conducting and analyzing multifactor experimental designs is fairly technical, so researchers should collaborate with experts to maximize the usefulness of the approach.<sup>1</sup>

## Limitations of Multifactor Experimental Designs

As noted, multifactor experimental designs require relatively homogenous experimental units. For example, with physician practices implementing alternative approaches, multifactor experimental designs assume that the practices have relatively homogenous outcomes prior to the study. Generally, the effects of differences between practices in outcomes due to patient characteristics can be controlled with a regression model that includes covariates for the measured patient characteristics. However, to avoid biased study results, any outlier practices with *vastly* different outcomes than the rest of the group before the intervention should be excluded from the study before the random assignment of treatment combinations. The goal is to avoid the situation in which pre-existing, inherent differences between an outlier and the other units produce differences in outcomes during the study period that cannot be distinguished from the experimental factors being tested.

If many components are being tested, it is possible that some significant findings resulted from chance, and that could possibly lead to overly optimistic conclusions about whether true effects exist. This problem can be addressed by conducting an overall F-test of whether the difference between alternative  $a$  and alternative  $b$  is zero for all components. If this test is rejected, the evaluator can use the t-tests to draw inferences about which components have meaningful effects. Researchers can also conduct joint tests for significance of related components.

In clustered designs in which practices implement the alternative approaches but patient-level outcomes are analyzed, statistical power depends primarily on the number of practices and less on the number of patients involved. If relatively few practices participate, the power to detect moderate main effects and interactions will be low. This limitation exists regardless of how many components are being tested. Without adequate power, statistically insignificant differences in outcomes between enhanced and routine versions of a care component cannot be taken as evidence that the routine (and typically less expensive) version of the intervention is just as effective as the enhanced version. Further, much larger sample sizes would be required to compare the effectiveness of one component to another, or to test interaction effects than would be needed for detecting main effects.<sup>13</sup> Another limitation in clustered designs in which practices implement the alternatives is that, if comparisons to routine care are made, it may be difficult to define routine care if practices are unaffiliated or dissimilar.

## Example of an Efficient Multifactor Experimental Design Study

To introduce health researchers to multifactorial experiments and to illustrate features of multifactorial experiments, we briefly describe a recently completed study that uses this methodology. Our study assessed the relative effectiveness of alternative ways of implementing various components of care management in special needs plans that serve disabled and frail elderly patients enrolled in both Medicare and Medicaid. One part of this study involved 24 care management teams at two plans implementing either routine

**Table 1. Assignments of Alternatives That Test Ways of Operationalizing Care Management in a Study at Special Needs Plans**

Care manager	Frequency of routine contact between care manager and member	Involvement of a medical nurse in management of complex medical cases	Follow-up during hospital admission and after discharge	Brown bag review of medication *
1	a) Contact frequency based on member risk	b) Medical nurse is always involved	a) Current practice: care manager contacts member during the admission, conducts an in-person follow-up at discharge, and monitors as needed	a) No brown bag review of medication
2	b) More frequent contact (also based on member risk)	a) A medical nurse is involved as needed	b) Current practice, plus additional follow-up within a week of discharge, plus monitoring	b) Care manager performs a brown bag review for members with 4+ prescriptions
3	a) Contact frequency based on member risk	b) Medical nurse is always involved	b) Current practice, plus additional follow-up within a week of discharge, plus monitoring	a) No brown bag review of medication
4	b) More frequent contact (also based on member risk)	a) A medical nurse is involved as needed	a) Current practice: care manager contacts member during the admission, conducts an in-person follow-up at discharge, and monitors as needed	b) Care manager performs a brown bag review for members with 4+ prescriptions

or enhanced care for each of 11 intervention components for a one-year period. Health outcomes assessed were (1) the number of inpatient admissions, (2) the incidence of readmission within 30 days of any discharge and within 30 days of a discharge for medical reasons, and (3) the number of emergency department visits.

The components we examined included (among others) alternative ways of conducting routine contacts with patients, screening for the risk of falls, depression screening, patient education, and management of care transitions. For example, for the components relating to care transitions, we tested the effectiveness of the plan's current practice of conducting one follow-up with patients after discharge from an inpatient setting versus an enhanced option that adds a second follow-up within a week of discharge. Table 1 illustrates 5 of 11 interventions studied. It shows care managers' assignments to demonstrate that each care manager implements a combination of alternatives *a* and *b*. In the interest of brevity, assignments for care managers 5–24 and interventions 5–11 are not shown.<sup>2</sup>

We learned from participating plans that an efficient multifactor experiment provided structure and consistency of expectations about intervention components, which can help standardize the intervention, leading to less variation in implementation across managers. This study also encouraged organizations to create a culture of learning and continuously seek ways to improve the components of their programs, while providing them with a

rigorous approach for testing out new ideas. A fidelity analysis, for which participants collected detailed data about implementation, enabled participating plans to learn about *how* each component is implemented relative to the study protocol, which can help identify which areas of care management should receive additional focus in their quality improvement efforts.

### Overview of EHR System Features

EHR systems, especially more advanced or robust systems, can serve as sources for rich data to inform ongoing care-improvement efforts and comparative effectiveness research as well as tools for enrollment and randomization when paired with multifactor experimental designs. Other than employing costly chart abstraction, well-designed and well-implemented EHR systems are the only single data source that can generate rich point-of-care health and service-provision data for monitoring and research purposes.

The type of data EHRs can generate depends on the type of EHR implemented, available data querying functionalities of the record system, and decisions made during implementation about which data to capture in structured data fields. Some basic EHR systems may primarily serve as documentation to support billing charges; more advanced EHR systems may include results management, order entry, decision support, and population health management.<sup>14</sup> In settings that have successfully implemented the more advanced EHR systems, structured data will likely be available for (1) key laboratory indicators, such as HbA1c, cholesterol level,

Note:\* In "brown bag" reviews, patients bring in all prescription and nonprescription medications.

and urinalysis; (2) patient characteristics, such as chronic conditions, insurance coverage, family history, smoking history, and risk factors; (3) clinical interventions that have been delivered, such as inpatient and outpatient procedures, prescribed medications, and patient education; and (4) information about service provision, such as frequency and dates of visits, missed appointments, and continuity of care at the clinician level. EHR systems with an integrated patient portal (Project HealthDesign,<sup>15</sup> for example) may also include patient-reported data on symptoms and quality of life.

In addition to offering greater opportunities to capture structured data, the more advanced EHR systems offer greater opportunities for making changes to computer-supported clinical work processes (such as providing data on patients' medication adherence to practice staff tasked with following up with those who have poorly controlled chronic conditions). Further, EHR systems can be used to support activities of randomized experiments, including enrollment and randomization, as recently illustrated by the Veterans Affairs study comparing two regimens of administering insulin.<sup>16</sup> To best support quality improvement and comparative effectiveness efforts that aim to answer new research questions (such as effectiveness of components of care), flexible and extendable strategies are needed in data collection, data access, decision support, and analytics.<sup>17</sup>

### Limitations of Using EHR Data for Studies

Despite the previously mentioned benefits of processing and extracting data from EHR systems, EHR vendors continue to place a low priority on developing more advanced EHR functions. The potential of this technology to capture, store, and report rich clinical data notwithstanding, EHR data currently suffer many limitations: incompleteness; a structure appropriate for management and billing processes rather than for monitoring, improvement, and comparative effectiveness research; and cumbersome processes for extracting the data.

These limitations arise from (1) lack of standardization of key data elements for information exchange with other systems within and between provider practices; (2) documentation challenges, such as the use of free text for assessment and care plans; (3) the rigidity of templates and other data fields for documenting and compiling information from within and outside the provider practice; (4) the constraints imposed by fee-for-service reimbursement, which encourages EHR use for documentation of billable events; (5) the tradeoff between capturing clinical data and using them for continuous quality improvement; and (6) missing data elements because the systems do not require them in templates and other structured documentation screens.<sup>18-20</sup>

Because EHR use by ambulatory care providers is still low (estimated as 20.5 percent using a basic system; 6.9 percent using a fully functional system in 2009,<sup>21</sup> it is unclear when the full promise of rich EHR data will materialize. For this reason, it is difficult to predict when a critical mass of providers using EHR systems will allow the construction of data files for monitoring and research.

Heterogeneity in the delivery of interventions can also pose a significant threat to valid inference. This threat should be understood not as a consequence of the use of efficient multifactor designs or of the use of EHRs themselves. Rather, it is a problem inherent in conducting experiments during the course of providing everyday patient care. Randomized clinical trials impose fairly strict controls on all aspects of a study subject's care and typically define precise criteria for inclusion in the study. The control and precision help ensure that differences in outcomes experienced by treatment and control subjects can be ascribed to the intervention. As useful as such control is, it is unavailable to studies ancillary to the provision of routine care. Nonetheless, the effects of heterogeneity can be managed reasonably well in most circumstances. Studies conducted within a health care practice or health system may not—as a result of internal guidelines and staff training—encounter significant heterogeneity. Periodic monitoring of and communication among providers may generally be sufficient to ensure essential equivalence in the delivery of care. In addition, preliminary analysis of data may suggest that interventions diverged from their expected pattern for particular providers or patients. This could lead to further investigation to ascertain whether the expected protocol was implemented as planned or whether some cases should be dropped from the analysis.

### Examples of Potential Uses of Efficient Multifactor Experimental Design and EHR Data for Quality Improvement

Multifactor experimental design studies are most appropriate in settings in which there is a willingness and ability to change routine approaches and to randomize implementation of tested alternatives. Because the power of multifactor experimental design lies in simultaneously testing the effectiveness of alternative ways of implementing several care components, participants must be willing to change their approaches to multiple components. For example, with physician practices as study participants, involving practices and individual clinicians in the selection of components to test can help increase their willingness to participate. Further, providing implementation guides and individualized assignment sheets that outline combinations of alternatives can prevent confusion and can also ease concerns about the difficulty of implementing the assigned combination of alternatives. Nonetheless, the success of a multifactor experiment relies on practices implementing assigned alternatives with high fidelity to the study protocol.

In this section, we give two examples—evaluating CDS features and studying components of a PCMH model—of how efficient designs can be used with EHR data for quality improvement. We conclude by providing additional examples of how this design and data can be used to evaluate the relative effectiveness of primary care characteristics typically measured in surveys of primary care practices.

## An Evaluation of CDS Features

Although CDS alerts and reminders within an EHR hold great promise for improving the safety and quality of patient care, especially for patients with chronic illnesses,<sup>22, 23</sup> some evidence suggests that CDS is often not well integrated into clinical workflow.<sup>24</sup> Because frequent electronic reminders can disrupt workflow and physicians may bypass or fail to respond to them,<sup>25-27</sup> it is important not only to test the benefit of CDS functions but also to understand mechanisms through which CDS alerts affect workflow. Because EHRs offer many CDS functionalities, and there are many ways to implement each functionality (e.g., in terms of triggers, timing, and type of staff alerted), efficient multifactor experimental design is a quick and rigorous way to simultaneously study many CDS components. EHR systems can be used to extract and record information on outcomes and to assess fidelity to process-of-care guidelines.

For example, one could test whether or how often to display alerts or reminders about evidence-based guidelines, as well as to whom to display them (e.g., to nurses or physicians). Medication management interventions could be tested—using the EHR to generate lists of patients with a potential interaction, contraindication, or out-of-range dosage, for example. Providers could then follow up with these patients through routine medication reconciliation procedures, or “brown bag” reviews, in which patients bring in all prescription and nonprescription medications. Multifactor experimental design allows both the reminder and the brown bag components of the medication management intervention to be assessed separately or together, and it also has the potential to identify interactions between the two.

Use of a diabetes care dashboard an important CDS function has been shown to increase the efficiency with which clinicians can access key clinical information necessary for the care of diabetes patients,<sup>28</sup> but additional studies are needed to test varying ways of seamlessly integrating such a data display into clinical work processes. A multicomponent study of this intervention could simultaneously assess: different graphic displays of information; work processes, including nursing staff access to the display compared to physician-only access; and differing methods of following up with patients based on clinician recommendations derived from review of the data. An efficient multifactor experimental design for such a comparative effectiveness study would improve the value of research by identifying which of the key changes in work process are the most important to emphasize in implementation efforts. Clinical parameters captured in the EHR, such as blood pressure, hemoglobin A1c, and LDL cholesterol levels, would provide appropriate outcome measures. Because these are collected in the course of regular diabetes care, the EHR could efficiently be part of the intervention and the means for assessing its effect.

## An Evaluation of Components of the PCMH Model

The PCMH is a promising model of care delivery that aims to strengthen the primary care foundation of the health care system by reorganizing the way primary care practices provide care. This reorganization includes the use of health information technology (health IT) for internal processes and for connecting the practice with its patients and with other providers.<sup>29</sup> Multifactor experimental designs can be used to study quality improvement components in general, but also to study the specific features of the PCMH.

The PCMH model encompasses five principles: (1) a patient-centered orientation; (2) comprehensive, team-based care; (3) coordinated care; (4) continuous access to care; and (5) a systems-based approach to quality and safety. One particular reason for considering using an efficient multifactor experimental design would be to explore how clinicians use EHR functionalities to deliver care that fulfills these principles. Of particular operational and research interest are such care coordination activities as (1) maintaining patient continuity with primary care physicians or primary care teams, (2) documenting and compiling patients’ information within and outside the primary care office, (3) sharing care with clinicians across practices and settings, and (4) managing care transitions.<sup>20</sup> Because the EHR system could generate information about the provider of care, the type of care, and the outcomes and other characteristics for all patients with complex medical cases in a PCMH, the assessment of the relative effectiveness of these components is highly feasible.<sup>3</sup>

## Other Examples

In addition to the uses summarized above, efficient designs and EHR data could also be used to examine the relative effectiveness of key components of primary care (many of which are commonly measured in surveys of primary care practices). Examples of these practice features include (1) access to and continuity of care, (2) planned chronic and preventive care, (3) risk-stratified care management, (4) patient and caregiver engagement, (5) care coordination across the medical neighborhood, and (6) continuous quality improvement.

## Conclusions

The use of efficient multifactor experimental designs, especially coupled with EHR data, has great promise for helping physician practices, hospitals, and other stakeholders conduct rapid-cycle comparative effectiveness research in order to assess alternative ways of implementing various aspects of care. It can also help improve efficiency by assessing whether more expensive approaches to a particular component of care yield better outcomes than “usual care.” Its power is its ability to enable scientifically rigorous testing of many facets of care provision simultaneously in real-world settings where change is ongoing.

As detailed above, many possible studies can be done, including testing alternative ways of implementing such critical primary components as screening patients for risk of specific adverse events (e.g., falls and depression), increasing patients' engagement in their care or adherence to evidence-based recommendations, coordinating the care of complex patients, improving prescription of medications, and encouraging various desired behaviors in clinicians. Efficient designs and EHR data can also be used to test how to most effectively use the EHR, by testing alternative dashboards, reminder systems, and clinical work processes. EHR data greatly facilitates multifactor experiments by providing timely data on clinical indicators and process-of-care measures as well as utilization and outcome data. It also enables assessment of fidelity to the assigned alternatives because EHRs offer a platform to extract and to record information about fidelity to process-of-care guidelines.

Effective use of efficient design for comparative effectiveness research is likely to require engagement of a researcher with expertise in this methodology, at least at the outset of the study. Further, successful implementation of multifactor experiments requires the engagement and buy-in of staff at all levels that will be affected, especially nurses or other staff who will be implementing the practice alternatives being tested. The study should not include overzealous monitoring to make certain that the assigned interventions are implemented as that level of oversight would not exist in an ongoing program. Interventions should be carefully chosen to include only those for which actionable changes will be made if the results of the study indicate that they are beneficial. Having a substantial number (e.g., 40 or more) of staff (or experimental units such as practice locations) participating in the study will yield much greater learning, as that would provide greater statistical precision for each care component tested. Dozens of care components can be tested simultaneously with this number of participating units.

Experimenters should beware of some of the pitfalls of EHR data: incompleteness, a structure appropriate for management and billing processes rather than for monitoring and improvement, and a costly and cumbersome process for extracting the data from EHR systems. Even though some of these limitations are similar to those inherent in health care administrative data, EHR data have the potential to overcome these challenges if incentives are set in place to standardize reporting and encourage use of the data for clinical and monitoring purposes.

An invaluable byproduct of multifactor experiments is the creation of a culture of continuous learning and improvement among staff at all levels—and across the health care delivery system as a whole. The increasing use of EHRs in both hospital and ambulatory care settings provides an opportunity to further leverage this methodology to rapidly identify effective quality improvements.

## Acknowledgements

We gratefully acknowledge the help of our colleague Arnold Chen in the preparation of this paper, as well as research assistance of Joe Brown.

## References

1. Fisher RA. The design of experiments. 1st ed. Edinburgh: Oliver and Boyd; 1935.
2. Box GE, Hunter SJ, Hunter WG. Statistics for experimenters: design, innovation, and discovery. 2nd ed. Hoboken, NJ: John Wiley & Sons; 2005. 664 p.
3. Zurovac J, Peikes D, Zutshi A, Brown R (Mathematica Policy Research, Washington, DC). Efficient orthogonal designs: testing the comparative effectiveness of alternative ways of implementing patient-centered medical home components. Rockville (MD): Agency for Healthcare Research and Quality; 2013. AHRQ Publication No. 13-0024-EF.
4. Peikes D, Zutshi A, Genevro J, Smith K, Parchman M, Meyers D (Mathematica Policy Research, Princeton, NJ). Early evidence on the patient-centered medical home. Final report. Rockville (MD): Agency for Healthcare Research and Quality; February 2012. AHRQ Publication No. 12-0020-EF.
5. Zurovac J, Brown R. Orthogonal design: a powerful method for comparative effectiveness research with multiple interventions. Washington (DC): Center on Health Care Effectiveness, Mathematica Policy Research; 2012. Document No. PP12-30.
6. Ledolter J, Swersey AJ. Testing 1-2-3: experimental design with applications in marketing and service operations. Stanford, CA: Stanford University Press; 2007.
7. Moore CH. Experimental design in health care. *Qual Manag Health Care*. 1994;2(2):1-15.
8. Jones FG, Moore CH. Designing and executing experiments in care—a data-driven, scientific approach to quality improvement. In: Schoenbaum SC, editor. *Measuring clinical care: a guide for physician executives*. Providence, RI: American College of Physician Executives; 1995.
9. Collins LM, Murphy SA, Strecher VJ. The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): new methods for more potent eHealth interventions. *Am J Prev Med*. 2007; 32:S112-8.
10. Collins LM, Baker TB, Mermelstein RJ, Piper ME, Jorenby DE, Smith SS, Christiansen BA, Schlam TR, Cook JW, Fiore MC. The multiphase optimization strategy for engineering effective tobacco use interventions. *Ann Behav Med*. 2001;41(2):208-26.
11. JMP [Internet]. Cary (NC): JMP; c2013. JMP: Powerful statistical analysis linked with interactive graphics, in memory and on the desktop; [cited 2013 Aug 30]; [about 17 screens]. Available from: <http://www.jmp.com/software/jmp10/>.
12. Minitab [Internet]. State College (PA): Minitab, Inc.; c2013. Minitab Statistical Software; [cited 2013 Dec 17]; [about 2 screens]. Available from: <http://www.minitab.com/en-US/products/minitab>.

13. Dziak JJ, Nahum-Shani I, Collins LM. Multilevel factorial experiments for developing behavioral interventions: power, sample size, and resource considerations. *Psychol Methods*. 2012;17(2):153–75.
14. Blumenthal D, Glaser JP. Information technology comes to medicine. *N Engl J Med*. 2007;356(24):2527–34.
15. Project HealthDesign [Internet]. Madison (WI): University of Wisconsin-Madison School of Nursing; c2012. Observations of daily living; [cited 2013 June 1]; [about 2 screens]. Available from: <http://www.projecthealthdesign.org/resources/observations-of-daily-living>.
16. D'Avolio L, Ferguson R, Goryachev S, Woods P, Sabin T, O'Neil J, Conrad C, Gillion J, Escalera J, Brophy M, Lavori P, Fiore L. Implementation of the Department of Veterans Affairs' first point-of-care clinical trial. *J Am Med Inform Assoc*. 2012;19:e170–6.
17. EDM Forum. Informatics tools and approaches to facilitate the use of electronic data for CER, PCOR, and QI: resources developed by the PROSPECT, DRN, and enhanced registry projects issue briefs and reports. Washington (DC): EDM Forum; 2013. Paper No. 11.
18. Fernandopulle R, Patel N. How the electronic health record did not measure up to the demands of our medical home practice. *Health Aff*. 2010;29(4):622–8.
19. Mandl KD, Kohane IS. Escaping the EHR trap—the future of health IT. *N Engl J Med*. 2012;366(24):2240–2.
20. O'Malley AS, Cohen GR, Kemper NM, Pham HH. Are electronic medical records helpful for care coordination? experiences of physician practices. *J Gen Intern Med*. 2010;25(3):177–85.
21. HealthIT.gov [Internet]. Washington (DC): U.S. Department of Health and Human Services. Policy making, regulation, & strategy: meaningful use; [cited 2013 May 2]; [about 3 screens]. Available from: <http://www.healthit.gov/policy-researchers-implementers/meaningful-use>.
22. Baron RJ. Quality improvement with an electronic health record: achievable, but not automatic. *Ann Intern Med*. 2007;147(8):549–52.
23. Garg AX, Adhikari NK, McDonald H, Rosas-Arellano MP, Devereaux PJ, Beyene J, Sam J, and Haynes RB. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes. *JAMA*. 2005;293(10):1223–38.
24. Carayon P, Karsh B, Cartmill R, Hoonakker P, Hundt AS, Krueger D, Thuemling T, Wetterneck T, Brennan PF, Walker J (Center for Quality and Productivity Improvement, University of Wisconsin-Madison, Madison, WI). Incorporating health IT into workflow redesign: request for information summary report. Rockville (MD): Agency for Healthcare Research and Quality; July 2010. AHRQ Publication No. 10-0074-EF
25. Downs SM, Anand V, Dugan TM, Carroll AE. You can lead a horse to water: physicians' responses to clinical reminders. AMIA Annual Symposium; 2010 Nov 13–17; Washington, DC. Bethesda (MD): American Medical Informatics Association; 2010.
26. Kuilboer MM, van Wijk MA, Mosseveld M, van der Lei J. AsthmaCritic: issues in designing a noninquisitive critiquing system for daily practice. *J Am Med Inform Assoc*. 2003;10(5):419–24.
27. Kuperman GJ, Bobb A, Payne TH, Avery AJ, Gandhi TK, Burns G, Classen DC, Bates DW. Medication-related clinical decision support in computerized provider order entry systems: a review. *J Am Med Inform Assoc*. 2007;14(1):29–40.
28. Koopman RJ, Kochendorfer KM, Moore JL, Mehr DR, Wakefield DS, Yadamsuren B, Coberly JS, Kruse RL, Wakefield BJ, Belden JL. A diabetes dashboard and physician efficiency and accuracy in accessing data needed for high-quality diabetes care. *Ann Fam Med*. 2011;9(5):398–405.
29. Moreno L, Peikes D, Krilla A (Mathematica Policy Research, Washington, DC). Necessary but not sufficient: The HITECH Act and health information technology's potential to build medical homes. Rockville (MD): Agency for Healthcare Research and Quality; 2010. AHRQ Publication No. 10-0080-EF

## Endnotes

\* For those who wish to explore using multifactorial experiments, Zurovac et al.<sup>3</sup> and Zurovac and Brown<sup>5</sup> provide a brief overview of design methodology. Box and colleagues<sup>2</sup> provide an accessible book on methods and theoretical background. For applications to marketing and service operations, see Ledolter and Swersey.<sup>6</sup> For the discussion of use of these designs in the health field, see Moore<sup>7</sup> and Jones and Moore.<sup>8</sup> For behavioral health research applications, see two articles written by Collins and colleagues.<sup>9,10</sup> Some of the most common software available for creating and analyzing multifactorial experiments are JMP<sup>11</sup> and Minitab.<sup>12</sup>

\*\* The full assignment grid (not shown) reflects a Plackett-Burman design with foldover.