Ethical Oversight of Quality Improvement and the Research-QI Boundary: A New Common Rule Changes Little

BY JOSHUA ROLNICK, N. LANCE DOWNING, LISA SHIEH, PAUL HEIDENREICH, AND MILDRED K. CHO

Quality improvement (QI) is now considered a fundamental part of the health care delivery system, and health care systems have an obligation to engage in QI activities. Yet these activities increasingly use methods employed in clinical research studies, such as treatment randomization, that contribute to the blurring of the distinction between clinical research and QI. An ongoing source of debate is the question of how to determine the boundary between QI and research for purposes of the federal human research regulations known as the Common Rule. When does evaluation of QI activities require review by an institutional review board (IRB)? When QI is not overseen by an IRB, what should institutions do to ensure that it is performed ethically?

The major changes to the Common Rule that the U.S. Department of Health and Human Services (HHS) and other federal agencies announced on January 18, 2017, fail to resolve those questions. Earlier proposed changes in the Notice of Proposed Rulemaking (NPRM) had included a new exclusion from the Common Rule requirements for "quality improvement activities," defined as interventions intended to change use of an accepted practice. The exclusion reflected a desire to create a new framework that "adequately accounts for the needs of a ‘learning’ health care system for continual quality improvement." The final rule discards this proposal, leaving largely intact how the original Common Rule treated QI, except for a new exemption affecting secondary analysis of patient information for QI.

We believe that HHS was correct to reject in the final rule the proposed QI exclusion in the NPRM, which provided a flawed attempt to distinguish QI from research. However, the final rule carries forward the problems of the original Common Rule for oversight of QI activities. First, uncertainty regarding the need for IRB review discourages careful efforts to understand the impact of individual QI efforts, exerting...
a chilling effect on the conduct and publication of QI evaluation, when ethical oversight of QI should result in incentives to improve health care delivery. Second, reliance on the Common Rule creates an inverse relationship between methodological rigor and oversight: systematic evaluation of QI interventions receives the administrative oversight of the IRB, while less systematic QI efforts are usually subject to little or no ethical review. We propose locating ethical oversight of QI outside the IRB system, in a different system that better integrates ethical oversight with operations, a change that would recognize the close connection between QI and clinical care.

To illustrate the problems with both the NPRM approach and the approach in the final Common Rule change, we use an example from our own QI experience: a randomized evaluation of a sepsis alert embedded in an electronic health records (EHR) system. EHRs have exerted a profound impact on QI and health systems reform. Yet the impact of these interventions is often difficult to disentangle from other concurrent QI activities, and even computerized alerts can cause harm. Randomization through the EHR is now a technological possibility, one with potential to become regular practice. Indeed, some have called for routine point-of-care randomization. However, there is no consensus on when randomization of an EHR-intervention requires IRB oversight, and randomization is sometimes cited as a sufficient condition to require IRB review.

Other practitioners of EHR randomization have reached contrary conclusions, and different practices are reported in the published literature. Some authors report having obtained IRB approval, others describe their evaluation activities as QI not requiring IRB review, and still others do not mention the IRB at all. Randomization of an EHR-based intervention thus provides an important example of the evolving boundary between QI evaluation and research. Our discussion, however, offers broader lessons for oversight of QI evaluation and the role of the Common Rule framework.

**Evaluation of a Computerized Sepsis Alert**

The debate over randomization of a sepsis alert took place within the context of a larger initiative at our hospital to improve sepsis outcomes. Here we describe that initiative. In 2013, our hospital reviewed its sepsis performance. We are part of a consortium of academic hospitals sharing sepsis outcomes, and each quarter, the consortium compares information on adjusted sepsis mortality. Although we sought to place in the top quartile in terms of adjusted mortality rate, we found that our actual performance had been closer to the median. We also reviewed process outcomes: how often our patients receive blood cultures, a serum lactate level, antibiotics, and intravenous fluids when indicated. Timely administration of intravenous antibiotics, in particular, improves outcomes, with mortality increasing for every hour antibiotics are delayed. We were focused on our rate of administering antibiotics within three hours of the onset of sepsis. We began a multipronged initiative to improve our treatment of sepsis, including increased education for clinicians, implementation of a nursing screening checklist, and special order sets.

The sepsis alert was conceived as one strategy to improve diagnosis and treatment. The alert focuses on patients with severe sepsis, defined as sepsis accompanied by low blood pressure or organ dysfunction. Embedded directly into the EHR system, an algorithm analyzes discrete data from the EHR to determine, in real time, if a patient has developed severe sepsis. The alert was designed to supplement, not supplant, other institutional initiatives to improve sepsis outcomes.

Although every new clinical decision support (CDS) tool is evaluated by a CDS committee prior to launch, the institution had never used randomization for evaluation. Standard practice would be to conduct a pilot trial of the alert in a subset of hospital wards for validation and then launch the alert across the institution. A before-and-after comparison might be used to evaluate its impact. These methods, however, could not disentangle the impact of the alert from other components of the sepsis initiative. Our EHR system provided the ability to randomize the alert by patient, offering an option to use randomization to help isolate the effect of the alert. The effect was difficult to predict. Although computerized sepsis alerts have become commonplace, their effect remains unclear, with some studies showing no impact on delivery of antibiotics or sepsis outcomes. More generally, the effects of using EHR-based alerts are uncertain. These CDS tools include drug interaction alerts, structured order sets, and other clinical reminders. When properly implemented, CDS tools can improve physician workflow, reduce drug administration times, and improve adherence to national guidelines. Yet CDS, like any medical intervention, can also potentially cause harm. Past studies have found, in the most extreme case, that a computerized order entry system in a pediatric hospital was associated with increased mortality.

We developed a plan for randomization. All patients admitted to wards and intermediate intensive care units (ICUs)—in other words, those who would be subject to the alert under a normal launch—would be randomized on admission to an active (treatment) or silent (control) group. In the treatment group, the alert would send an automated page to a covering physician and a
specially trained crisis nurse, who would assess the patient. In the control group, no page would be generated. We performed all the usual steps in a randomized evaluation: we defined primary and secondary endpoints, conducted a power analysis, and scheduled an interim assessment. In planning our evaluation, however, we faced important ethical questions. Is it ethical to randomize this new sepsis alert? If we randomize, are we engaged in human subjects research requiring IRB approval and perhaps necessitating patient consent? To help answer these questions, we obtained a research ethics consultation. We then describe how we resolved questions for our sepsis alert and offer general lessons for ethical oversight of QI activities.

**Definition of Quality Improvement**

The first step in understanding how to apply the Common Rule is to define QI. While neither the original version of the Common Rule nor the final regulatory rule change defines QI, the published literature contains several helpful attempts at a definition.

The literature has used two approaches. The first is to offer a conceptual definition of QI. The second is to identify a series of attributes that separate QI from research along a QI-research continuum. Batalden and Davidoff defined quality improvement broadly as the "combined and unceasing efforts of everyone—health care professionals, patients and their families, researchers, payers, planners and educators—to make the changes that will lead to better patient outcomes (health), better system performance (care) and better professional development." Their definition encompasses the vast, heterogeneous set of activities that may fall under the umbrella of QI, including computerized sepsis alerts, new nursing checklists, and even the creation of a quality officer to oversee institutional systems improvement. It also would make medical research, both basic and clinical, a subset of QI. In contrast, Lynn et al., following Baily et al., defined QI more narrowly as "systematic, data-guided activities designed to bring about immediate improvements in health care delivery in particular settings." This definition emphasizes systematic evaluation as an essential part of QI. The definition does not directly address if, or when, QI and research overlap.

Ogrinc et al. developed and validated an instrument for the purpose of distinguishing QI from clinical research. Their framework recognizes that clinical research and QI operate on a continuum. The framework identifies six domains: (1) intent and background, (2) methods, (3) intended benefit, (4) risk, (5) applicability of results, and (6) sharing and disseminating. QI has several characteristic qualities: the background and intended benefit are local to the institution and its patients, statistical methods measure systems-level outcomes over time rather than compare outcomes between groups, risks of nonparticipation are generally considered higher than risks of participation, and results are applied and disseminated first at the local, institutional level. Ogrinc et al.'s instrument is intended to help IRBs distinguish between QI and research. It is less clear that the framework is intended to be prescriptive of what QI should be. For example, the instrument describes statistical methods of QI as tending to "evaluate system level processes and outcomes over time with statistical process control or other methods," whereas the methods of research "primarily compare differences between groups or correlate observed differences with a known health condition." Yet, as we describe, treatment randomization can be used in activities that resemble QI.

Both approaches have strengths. Lynn et al. identifies an important feature of QI—a primary focus on bringing about immediate improvements in particular settings—although QI can have secondary objectives as well. Ogrinc is right to note that QI and research occupy a continuum, and addressing the imperfect distinctions between QI and research is essential to understanding more about the ethics and law of QI oversight. No definition of QI can sharply demarcate its borders, but when these definitions are integrated, several important features of QI activities are suggested. First, the primary, but not only, focus is immediate and local, to bring out immediate improvements in particular settings. Second, QI activities use interventions that appear ex ante to pose little incremental risk over usual care, although in individual cases, the ex post impact may be significant. Third, the interventions of QI tend to use modifications to organizational practice, including changes to training, information technology, and staffing protocols, rather than the introduction of new medications or invasive procedures. Finally, QI is closely connected to the enterprise of routine clinical care.

It is also useful to separate two aspects of QI activities: development of QI interventions and evaluation...
of QI interventions. Changes to systems can occur with the intention of bringing about improvements in health care delivery, yet the impact of these changes may or may not be measured. For example, a sepsis initiative could include a computerized alert designed to improve sepsis outcomes in a particular health care setting, but the institution may never plan to measure the alert’s impact. Alternatively, the institution might use a variety of methods and types of information—such as treatment randomization and time series data—to measure the alert’s impact. It is when a QI intervention is subject to systematic evaluation that the question of IRB review arises.

QI evaluation has its own ethical principles that we shall explore in the next section.

**Ethical Principles of Quality Improvement Activities**

What are the ethical principles that should guide the development and evaluation of QI interventions? Lynn et al. list seven requirements of ethical QI: social or scientific value, scientific validity, fair participant selection, favorable risk-benefit ratio, respect for participants, informed consent, and independent review. The framework emphasizes some similar themes to research ethics: minimize risk to the extent possible, protect participants and their information, and ensure that QI is held accountable through appropriate oversight. Similarities stem from the common appeal to several principles of ethical conduct in health care, including respect for autonomy, beneficence, nonmaleficence, and justice.\(^5\)

Yet while there are some similarities, there are also some differences. These differences flow from the close connection between QI and routine medical care and the use of interventions posing minimal incremental risk over the baseline state. As Lynn et al. recognize, “Quality improvement is an intrinsic part of good clinical care, in which data from clinicians’ own settings guide them in improving their practices.”\(^26\) Batalden et al. argue that it is even possible to view health care as a service coproduced by patients and health care professionals, with patients possessing rights as well as some circumscribed obligations of participation.\(^27\) This contrasts with clinical research, in which participant-patients are understood to have the right to withdraw at any time.\(^28\) Because of the close connection to routine care, patients consent to participate in the QI process, at least to a reasonable degree, by consenting to medical care. While research carries a default presumption of explicit informed consent, subject to override, QI evaluation may appropriately carry a default presumption of implicit consent.

Obligations of participation also extend to health care professionals and health care systems. Evaluation should be as methodologically rigorous as other constraints permit, and results should be disseminated when possible. As Batalden et al. write, to know that a “change is producing improvement, we need accurate and powerful measurements of what is happening.”\(^29\) Because some QI interventions do not achieve their intended purpose, it is important to evaluate their impact before integrating an intervention into routine clinical practice.\(^30\) Methodological rigor is often subject to other constraints, of course. It will not always be possible to use the strongest available methods of evaluation, but it is important to attempt to do so, subject to practical constraints. While the primary objective is improvement in a specific setting, a secondary objective should be to improve health care delivery more generally. Sharing results with others serves a more general goal of health care delivery and promotes reciprocal dissemination that will benefit future QI activities.

As a corollary to using strong methods, QI practitioners should use randomization when appropriate. Randomization is well-suited to the needs of QI evaluation, especially interventions that occur through use of the EHR system. QI interventions tend to occur in clumps, under a more general objective to improve health care delivery, such as to lower sepsis mortality or reduce rates of postsurgical infections. Randomization is often the most effective technique to separate the effect of one QI intervention from another. These individual interventions often appear ex ante to pose minimal marginal risk while offering uncertain benefit. Of course, randomization is not always the most appropriate choice. Often, practical constraints foreclose its use. In other situations, strong ex ante conviction that an intervention improves on usual care may make randomization inappropriate. Further, EHR-based randomization makes randomization feasible in a wider range of situations, by lowering the resource barriers to randomization. The initial treatment assignment of EHR-based interventions is often staggered, based on nonclinical factors: one ward might receive an EHR alert prior to another because it happens to be chosen first by an informatics committee or to have a nurse manager interested in sepsis. Randomization offers an alternative way of staggering introduction, but one that will better establish cause and effect.

Finally, appropriate accountability is an ethical feature of QI evaluation. Just as health care systems and individual providers are held accountable for routine medical care, they should be held accountable for ethical conduct of QI. As Lynn et al. describe, accountability for the ethical conduct of QI should be integrated into practices that ensure accountability for clinical care, with a level of ethical review appropriate to the individual QI activity.\(^31\)
Oversight of QI and the Original Common Rule

Developed in the early 1980s, the Common Rule refers to the federal regulations governing most human subjects research in the United States. The federal rules were based on the findings of the Belmont Report, issued in 1979 by a national commission, which recognized core ethical principles of respect for persons, beneficence, and justice. The Common Rule set out a framework for research oversight, including institutional review, consent, and other measures. Since its creation, the Common Rule had changed little until recently.32

The original Common Rule does not define QI or state when evaluation of a QI intervention is covered by the regulation. Instead, the definition of research is determinative: an activity is excluded if it lies outside the definition of human subjects research.33 In both the original Common Rule and the new version pursuant to HHS's final rule, research is defined as "a systematic investigation, including research development, testing and evaluation, designed to develop or contribute to generalizable knowledge."34 Thus, it is the nature of the evaluation that determines if a QI initiative is considered research. It is not the type of intervention or the level of risk. Instead, evaluating a QI intervention makes the Common Rule applicable. Implementing a computerized alert for sepsis is not sufficient, but evaluating the alert may qualify as research.

Intended Generalizability. The most commonly discussed component of the Common Rule definition of research is the part that refers to an investigation "designed to develop or contribute to generalizable knowledge."35 No definition of "generalizable knowledge" is provided, and because IRB decisions are private, there is no body of prior IRB decisions to use as a guide for how IRBs have defined the term. The Common Rule appears to focus on the evaluation's intent; the evaluation must be "designed" to produce "generalizable" knowledge. Thus, if hospital A randomizes a sepsis alert for the sole purpose of evaluating how the alert would affect antibiotic administration in hospital A only—and not other institutions—the evaluation is not designed to generalize.

All investigations, however, are intended to generalize to an imperfect degree, and there is little to guide the process of determining if the threshold level of generalizability is satisfied in an individual case. All QI evaluations yield findings that generalize to some other settings but not to all. Yet the same holds for clinical research. Large randomized double-blind trials of new medications are often held up as the paradigmatic cases of investigations that produce generalizable knowledge, yet most randomized trials have long lists of exclusion criteria that limit generalizability to "real-world" settings by the study's design.36 IRBs themselves reach inconsistent decisions, making predictions difficult.37 Every practitioner of QI wishes for a QI evaluation to yield insights for other institutions but recognizes that institutional differences limit generalizability. Publication has been used as a proxy for intent, but QI practitioners may wish to publish their results even if results generalize only imperfectly. Conversely, many research studies are never published, and a study could be research even if it is never intended to be published.

Systematic Investigation. The "systematic investigation" requirement is also important because systematic evaluation of a QI intervention is what can prompt the question of IRB review. Once a systematic attempt is made to evaluate a computerized sepsis alert or a nursing checklist, then the question arises, is this evaluation designed to produce generalizable knowledge? QI evaluation should be systematic. At times, it may face practical constraints that a clinical research protocol might not, due to the close connection between QI activities and routine medical care. It may not be possible, for example, to use a nursing checklist across all nursing floors. Since the Common Rule does not define "systematic," there is little guidance on when QI evaluation becomes sufficiently systematic, again creating ambiguity.

Patient Consent. Although consent does not determine IRB applicability, the need for explicit patient consent is an important implication of the Common Rule. No investigator may involve a human being as a subject in research until the investigator has obtained "the legally effective informed consent of the subject or the subject's legally authorized representative."38 In both the old and new Common Rule, consent information must include "any reasonably foreseeable risks or discomforts," as well as "any benefits to the subjects or others."39 In contrast, health care facilities do not ask patients to explicitly consent to systems reform, such as a new institutional procedure to avoid medication errors.

There are substantial hurdles to obtaining patients' informed consent to initiatives traditionally conceived as QI. First, if a patient declines to participate in randomization, to what baseline does the patient return? If a patient requests not to participate in a sepsis alert evaluation, would the alert be disabled for the patient? Second, how would consent be obtained? The alert was intended for all hospital inpatients, only a small percentage of whom will develop sepsis. Would every patient sign a consent form on admission to the hospital? Given the large number and heterogeneity of patients entering the hospital via the emergency department, outpatient clinics, and hospital transfer, who would obtain the consent and how?
And could patients understand what it means to grant permission to be monitored by a computerized alert algorithm for presence of a condition that they do not have but could develop in the future?

While informed consent is required by default, exceptions in the old and new Common Rule are permitted when: 1) the research involves no more than minimal risk to the subjects; 2) the waiver or alteration will not adversely affect the rights and welfare of the subjects; 3) the research could not practicably be carried out without the waiver or alteration; and 4) the subjects will, whenever appropriate, be provided with additional pertinent information about participation. While the Common Rule might permit a waiver for EHR randomization in some circumstances, conditions for a waiver are unclear. What level of risk is “minimal?” By definition, minimal risk is present when “the probability and magnitude of harm or discomfort anticipated in the research are not greater in and of themselves than those ordinarily encountered in daily life or during the performance of routine physical or psychological examinations or tests.” In the context of a QI evaluation, however, a patient may face a baseline level of risk different from that in ordinary life. To what extent does the level of risk absent the intervention matter for the determination? IRBs have reached inconsistent conclusions. Second, the meanings of “rights and welfare” and “practically be carried out without the waiver” have been debated.

The NPRM’s New Exclusion for Quality Improvement

In September 2015, HHS released the NPRM proposing changes to the Common Rule and solicited comments about the proposed changes. Significant for QI, the proposal created a new category of QI activities that could be systematic and intended to generalize yet still excluded from IRB oversight. Evaluation was excluded as QI when the purposes were “limited to altering the utilization of the accepted practice and collecting data or biospecimens to evaluate the effects on the utilization of the practice.” Thus, the technique of evaluation was not determinative—a randomized evaluation could qualify. The rationale was that “[t]hese efforts, some of which could be judged to be research, should be carried out because of the recognized public good they achieve.”

The proposed changes in the NPRM thus shifted the focus from the method of evaluation (randomization, pre- or post-assessment, other) to the purpose of the intervention. Evaluation of an intervention would be excluded if the intervention were designed to promote utilization of an accepted practice rather than introduce a new practice. For example, a randomized evaluation of a handwashing initiative, such as an educational workshop for nurses, would be QI if the intervention is designed only to increase rates of handwashing, because handwashing is considered standard practice. Likewise, randomized evaluation of a checklist to prevent central-line infections would be QI because it is designed to promote adoption of practices known to prevent such infections. In contrast, randomized evaluation of a new method to prevent central-line infection would not qualify as QI. Nor would evaluation of whether any of the practices promoted by the checklist, such as using a mask, be part of accepted practice.

A problem the proposed changes in the NPRM did not acknowledge is that the boundary between changing utilization of accepted practice and changing practice is not always clear. Many CDS interventions straddle this line. For example, does a computerized sepsis alert promote adoption or change practice? If the alert is considered a method to promote completion of the sepsis bundle, then the alert is considered QI as a method to promote adoption of an accepted practice (the sepsis bundle). However, if the alert is viewed as a new test for sepsis, then a study would be categorized as research because the study is evaluating a practice. The same problem would accompany many other interventions to improve sepsis detection, such as a bedside nursing checklist. Many QI interventions occupy a place between influencing clinical practice and changing it.

Further, the NPRM suggested that evaluating the intervention should measure only changes in use of accepted practice, not the second-order effects on the outcome measures that the accepted practice is designed to influence. Evaluation of an educational workshop on handwashing, for example, should measure the effect on handwashing rates, not rates of nosocomial infection; evaluation of a sepsis alert should measure sepsis bundle completion, not mortality. Yet the purpose in changing accepted practice is, and should be, to improve patient outcomes, and it is not clear why a QI activity should be excluded from IRB oversight because the endpoints are process metrics for accepted treatment rather than outcome measures. The NPRM’s proposed change would have discouraged careful evaluation of QI by offering an incentive to limit evaluation to process measures that may not influence care: a sepsis awareness campaign might increase antibiotic utilization but in a fashion that did not improve sepsis outcomes.

The Final Revised Rule

Recognizing these problems, HHS eliminated the QI provision from the final rule, noting that it “could create more confusion than it resolved, and it might have
 inadvertently created inappropriate obstacles to those [quality assurance]/QI activities that should not fall under the rule. 52 However, instead of offering a revised definition, the final rule retains the framework of the original Common Rule. The same definition of research, with a focus on intended generalizability, continues to do the work of defining when the Common Rule requirements apply to QI activities.

The final rule also revises the original Common Rule’s requirements for the consent process but does not clarify the conditions for waiver. The consent changes focus on improving a subject’s understanding of a study’s purpose, risks, and benefits. 53 The final rule allows use of general consent for future use of identifiable private information and biospecimens. However, the requirements for waiver remain nearly the same, without clarification of terms. For example, commentary from HHS in the final rule notes, that “[a]fter considering the diversity of opinions expressed in the public comments on this issue, including many comments seeking further guidance concerning the proper interpretation of the ‘practically’ language, the final rule does not define this language (which was also included in the pre-2018 rule).” 54

A new exemption for secondary research with private identifiable information is one significant change for QI, but it too will add ambiguity and create incentives contrary to rigorous QI. The final rule adds several exemptions that, when applicable, eliminate most oversight requirements. One of these covers “[s]econdary research uses of identifiable private information or identifiable biospecimens” when the “research involves only information collection and analysis involving the investigator’s uses of identifiable health information” if information use is regulated as health care operations or research under the Health Insurance Portability and Accountability Act (HIPAA). 55 This exemption encompasses secondary uses of patient data regulated by HIPAA. However, data collection must have occurred for “some other ‘primary’ or ‘initial’ activity.” 54 Post-hoc QI evaluation with EHR data could fall under this exemption (although still subject to HIPAA requirements), but not an evaluation involving primary data collection.

For example, consider two different circumstances. In the first, a sepsis alert is introduced for patients on ward A but not ward B because the hospital decides it lacks the financial resources to introduce the alert on both wards simultaneously. Ward A is chosen arbitrarily. Data from both A and B could be used after the fact to analyze the alert’s effect to decide whether the hospital should permanently adopt the alert. In a second case, the hospital wishes to evaluate an alert and decides to randomize the alert to one ward but not the other, in order to have a comparator for evaluation. This evaluation would entail more than secondary use of data and appear not to be covered by the exemption. The exact contours of the exemption for QI, however, are unclear: what if there is more than one motivation for introducing the alert to only ward A?

The ambiguities in the new Common Rule, like the old, discourage ethical evaluation of QI. We have discussed the ethical obligation to evaluate QI interventions, using methodologically rigorous methods when institutional constraints permit. The Common Rule, however, applies only if the QI intervention is evaluated systematically. In the example of a computerized sepsis alert, the alert could be implemented without any form of systematic evaluation. A lack of evaluation might violate the ethical principles of QI, but it would resolve ambiguity over the application of the Common Rule and its attendant requirements. In addition, the Common Rule discourages techniques that may increase generalizability, such as evaluating a QI intervention’s impact on a broader patient population. Steps to increase generalizability would seem to enhance the scientific value of a QI evaluation. 55 Sharing results of QI evaluations should be encouraged, but under the Common Rule, dissemination may be taken as an indication that the evaluation was designed to produce generalizable knowledge.

Common Rule applicability also continues to prompt confusion over consent. As discussed, patients have special obligations to participate in QI activities. Implicit consent through participation in the health care system may be adequate in some situations. Yet confusion over when explicit consent can be waived may impede QI evaluation. How can individual participants withdraw from a CDS tool? A benefit of EHR randomization is the low-resource source of randomization, but requiring explicit consent obviates this advantage.

Our Decision about the Severe Sepsis Alert

Under the Common Rule, our randomized evaluation was systematic. But was it intended to produce generalizable knowledge? Here, it is important to note that randomization does not in itself render a study’s findings more generalizable. The use of randomization helps establish causality but does not improve the ability to generalize that causal link to other settings. We expected that our findings would have some generalizable lessons but not apply universally. Another institution with staff protocols and an organizational culture that are similar to ours may experience similar effects with a computerized alert. Sepsis alerts tend to produce similar results across institutions, improving process measures, such as those for intravenous fluids, but not sepsis.
mortality.56 However, generalizability would be limited. The impact depends to a great degree on how the alert is used. For our institution, the alert’s impact depended on dedicated intensive-care nursing resources and our specific clinical workflow to identify and treat sepsis. An institution with differences in staffing or culture might observe a different effect. More generally, the literature on CDS points to local context and implementation as key factors to success.57 We concluded that our randomized evaluation of the sepsis alert was not generalizable in the manner conceived by the rules because the primary intention was to inform local QI.

After the ethics consultation, we received a formal decision from our IRB that our randomized evaluation was not subject to IRB oversight. In addition to questions of generalizability, the ethics consultation focused on minimal level of incremental risk—prior evidence on sepsis alerts showed no clear evidence of an impact on care, and thus the level of incremental risk posed by the alert was minimal.58 Nevertheless, important questions were left unresolved: What level of intended generalizability and systematic design should trigger IRB oversight? If considered research, would our evaluation require patient consent? Is framing minimal risk in terms of incremental risk compared to the status quo consistent with the rules, even if usual care for hospitalized sepsis patients is greater than the risks of ordinary life? These considerations remain the same under the revised Common Rule.

Taking a step back, we observed that the Common Rule had discouraged rigorous evaluation of our QI intervention and discouraged use of EHR randomization more generally. The focus on generalizability appeared misplaced. Would it not be beneficial for our evaluation to yield lessons for other institutions? Why should intended generalizability trigger additional ethical oversight?

Should Ethical Oversight of QI Be Separate from Research?

How might we create a better system of ethical oversight for QI activities? How should the Common Rule have been revised? The system should be one that ensures appropriate accountability and ethical treatment of patients while encouraging ethical principles of rigorous QI evaluation, dissemination of results, and patient engagement in QI. One option is to broaden the Common Rule’s scope, clarifying that systematic QI is subject to its provisions, and tailor QI provisions of the Common Rule to QI. A second option is to create a separate system of oversight.

The Common Rule could be modified to clarify that all systematic QI is subject to oversight. QI activities could be defined in a manner highlighting the features we have discussed. For example, QI activities might be defined as the development and systematic evaluation of minimal risk changes to organizational practice for the purpose of improving health care delivery in specific settings. No definition is perfect, but an affirmative definition drawn from the literature would bring better clarity. Yet this approach—clarifying that the Common Rule applies to all systematic QI—would have shortcomings. First, it would preserve the incentive not to evaluate QI systematically, as a way to avoid IRB review. Rigorous evaluation of QI already faces institutional barriers, as it demands staffing time and resources, and IRB review would add another disincentive. Second, it would use a system that, once applicable, is poorly designed for the distinctive features of QI—the close connection between QI activities and patient care and the collaboration between patients and health care professionals. Consent requirements, for example, may not match the needs and ethical obligations of QI evaluation. While the final rule contains exceptions for consent, the criteria remain ambiguous. The exceptions reflect a strong presumption of explicit informed consent appropriate to clinical research but not to QI activities.

An alternative is to broaden and clarify the QI exclusion. The problem is that once QI exits the Common Rule, it enters an oversight vacuum. However, Common Rule modification could occur along with an effort across institutions to develop a parallel system of ethical oversight for QI. Some institutions have already developed such a system. Wise describes a system for review of QI evaluations meeting the federal definition of research (QI-IRB):

Our hospital has completed its first year of administering a QI-IRB following nearly all of the recommendations from the special report authored by Baily et al. The QI-IRB contains some of the same members of our research IRB, but it is distinguished by members whose expertise lies both in clinical research and QI. . . . We are part of a health care network, a Qualis (Medicare) network, and are bound to employ efficient processes, so we hope that some of our findings will be sufficiently generalizable that sharing with other organizations may contribute to improvements in clinical operations at other sites. Thus, once again, the overlap between our QI activities and the federal definition of human research looms over our activities. Providing a specialized ethics oversight body for QI studies has helped enrich our QI culture without inhibiting QI activities. QI study teams feel less isolated . . . .

Wise followed recommendations in Baily et al. to better integrate ethical review of human participants into accountability systems for clinical care.59 When QI qualifies as research (QI-research), the Common Rule can be accommodated by cre-
Fusing research over the ethical oversight of QI evaluation. In the present regulatory environment, there is sufficient rational to locate review of some efforts to evaluate CDS within such a committee. When a particular evaluation has sufficient intended generalizability to raise doubt, however, absent regulatory change, it must still pass through the IRB. In that case, the local IRB can be enlisted to help make a determination.

QI oversight can promote ethical evaluation in part by encouraging the use of careful techniques to measure a QI intervention’s effect. How this would be done may differ between different types of interventions. In the case of electronic CDS, for example, randomization may often be an attractive option to isolate effect. When feasible and appropriate, a CDS oversight committee should encourage use of methods like randomization, and one function of the committee would be to consider particular circumstances that make randomization more or less desirable. In other cases, the committee might decide that randomization is not justified or other methods are preferred. Additionally, QI oversight should promote dissemination of results so that other institutions can learn from each other’s experiences. Oversight would both protect patients and help ensure that QI activities are best designed to promote systems improvement.

Joshua Rolnick, MD, JD, is a clinical scholar in the National Clinician Scholars Program at the University of Pennsylvania; N. Lance Downing, MD, is a clinical assistant professor of medicine at Stanford University; Lisa Shieh, MD, PhD, is a clinical professor of medicine at Stanford University; Paul Heiderreich, MD, MS, is a professor of medicine and health policy and research at Stanford University; and Mildred K. Cho, PhD, is a professor of pediatrics and the associate director of the Center for Biomedical Ethics at Stanford University, which includes members with content expertise in QI. The committee could develop a subcommittee tasked with ethical oversight of QI evaluation. In the present regulatory environment, there is sufficient rationale to locate review of some efforts to evaluate CDS within such a committee. When a particular evaluation has sufficient intended generalizability to raise doubt, however, absent regulatory change, it must still pass through the IRB. In that case, the local IRB can be enlisted to help make a determination.

QI oversight can promote ethical evaluation in part by encouraging the use of careful techniques to measure a QI intervention’s effect. How this would be done may differ between different types of interventions. In the case of electronic CDS, for example, randomization may often be an attractive option to isolate effect. When feasible and appropriate, a CDS oversight committee should encourage use of methods like randomization, and one function of the committee would be to consider particular circumstances that make randomization more or less desirable. In other cases, the committee might decide that randomization is not justified or other methods are preferred. Additionally, QI oversight should promote dissemination of results so that other institutions can learn from each other’s experiences. Oversight would both protect patients and help ensure that QI activities are best designed to promote systems improvement.

Joshua Rolnick, MD, JD, is a clinical scholar in the National Clinician Scholars Program at the University of Pennsylvania; N. Lance Downing, MD, is a clinical assistant professor of medicine at Stanford University; Lisa Shieh, MD, PhD, is a clinical professor of medicine at Stanford University; Paul Heiderreich, MD, MS, is a professor of medicine and health policy and research at Stanford University; and Mildred K. Cho, PhD, is a professor of pediatrics and the associate director of the Center for Biomedical Ethics at Stanford University.

Acknowledgment

Mildred K. Cho was supported in part by the National Institutes of Health Clinical and Translational Science Award U54 RR024374-01 (with Harry Greenberg as the principal investigator).

References

5. 28 CFR 46.104(d)(4).
6. As of February 5, 2017, the final rule was scheduled to take effect in 2018. However, the Congressional Review Act allows Congress 60 days to pass a resolution of disapproval for presidential signatures. Two bills passed by the House, the Midnight Rules Relief Act and the Regulations from the Executive in Need of Scrutiny Act (REINS), would allow Congress alternative options to disapprove the Rule. Regardless of its fate, the final rule offers lessons for the challenges in applying the Common Rule framework to QI.
12. Angus DC. Fusing randomized trials


21. The research ethics consultation service is supported by a National Institutes of Health Clinical and Translational Sciences Award, U54 RR024574-01 (with Harry Greenberg as the principal investigator).


