Learning Health Systems and the Revised Common Rule

Joshua A. Rolnick

Introduction

Learning health systems (LHS) embed "knowledge generation ... into the core of the practice of medicine."[1] Their efforts to evaluate and improve the care they provide to their patients are quality improvement (QI) activities. LHS thus engage in continual quality improvement as part of their operations.[2] I assume here that LHS are a net positive development in health care delivery. I also assume that their QI activities are a desirable practice and public policy should promote QI. Subject to considerations such as cost, safety, and practicality, QI activities should be designed with rigorous internal methodology. That is, QI should have high internal validity: LHS must be able to infer the effects of changes in practice to understand how to make improvements. While internal validity is more important than external validity, because QI is primarily a local endeavor, QI should endeavor to offer lessons for other health systems. High external validity, the ability to generalize results to other settings, is therefore also a desired feature of QI at the margin, subject to other constraints.

Although the Common Rule (CR or "Rule") was created to regulate human subjects research, it has been interpreted to apply to some QI activities.[3] From a regulatory definitional point of view, some QI activities appear to also be human subjects research activities within the scope of the CR. Prior to the CR revisions, a great deal was written about the ambiguous boundary between QI and human subjects research.[4] My focus here is post-revision — where we are after the CR changes? Despite hope that the revised CR would provide a clearer QI-research boundary, the revised CR leaves the same ambiguities largely intact.[5] These ambiguities impose costs on the performance of "high-validity" QI. I shall sketch some guidance for health systems and regulators on operating within the revised Rule, and close with a suggestion for how to modify the Rule. I offer a proposed regulatory exclusion for QI activities, defined as an evaluation performed by a health system to improve the health system's care to its patients, provided such activities impose no risks greater than usual care.

Joshua A. Rolnick, M.D., J.D., M.S., *is affiliated with the Division of General Internal Medicine, Department of Medicine, Perelman School of Medicine, University of Pennsylvania, the National Clinician Scholars Program, University of Pennsylvania, the Palliative and Advanced Illness Research Center, Department of Medicine, Perelman School of Medicine, University of Pennsylvania, and the Corporal Michael J. Crescenz Philadelphia VA Medical Center.*

The Journal of Law, Medicine & Ethics, 47 (2019): 238-246. © 2019 The Author(s)
DOI: 10.1177/1073110519857279

https://doi.org/10.1177/1073110519857279 Published online by Cambridge University Press

Trends in Quality Improvement

American health care systems are experiencing an ongoing evolution, one in which the characteristics of QI are changing.[6] Three factors have converged to change QI practices. These forces have made QI more closely resemble human subjects research. First, health information technology has changed both the methodologies of QI and the outcomes that may be measured. Almost all U.S. health care systems have adopted electronic health record systems (EHRs).[7] Before EHRs, information for QI had to be abstracted from charts by hand, making large studies with detailed patient information impracticable. The same was true of research, but because research is more likely to collect information prospectively, the limitations of paper charts were especially concerning for QI. EHRs store information about patient characteristics, care utilization, and clinical outcomes in structured data fields. When information is not structured, it is increasingly extractable through natural language processing.[8] EHR information makes measuring clinical outcomes easier. For example, suppose a hospital takes steps to encourage physicians to wash their hands. Data from the EHR may be queried to measure rates of nosocomial infection.

In addition, EHRs have lowered the implementation costs of using randomization as a tool for evaluation. EHRs have produced a new class of QI tools, computerized early warning systems.[9] These systems use EHR data to look for a current or predicted future state, such as clinical deterioration, and when the state is found, notify clinical staff through automated pages and "pop up" alerts. Randomization is an emerging feature of EHRs.[10] Randomization through the EHR reduces the high expense and resource demands of randomized evaluation because the electronic health record can be used to randomize use of the warning system. Since some commercial EHRs allow such warning systems to be active or silenced in a randomized fashion, hospitals may have a practical option to evaluate the warning system in a randomized fashion.[11] For example, a hospital might develop a computerized early warning system to predict clinical deterioration and notify clinicians through a page to the attending physician of record. This warning system could be randomized to be active (i.e. send the notification) for a randomly chosen set of patients.

Second, changes in reimbursement have altered the incentives to perform QI well. The fee-for-service model of reimbursement offers no direct financial incentives to improve the quality of care — it may even offer disincentives for QI, at least when QI might reduce utilization of high-intensity services. While fee-for-service remains the most common form of reimbursement, alternate payment programs have become increasingly common.[12] For example, the Affordable Care Act created several programs to compensate hospitals and providers in alternate forms under which quality improvement may lead to higher reimbursement.[13] These programs have encouraged health systems to expend resources on QI initiatives.

Third, the consolidation of health systems has changed the generalizability of QI. Hospitals continue to merge at increasing rates.[14] Consolidation sometimes leads to merging of different types of hospitals, including mergers between academic medical centers and community hospitals, thereby increasing the heterogeneity within health systems.[15] Reduced competition may produce uncertain incentives for QI, but it also shifts the definition of "local" from a single hospital to a multi-institutional system, one that may have different types of care settings within it.[16] QI initiatives that span multiple hospitals in a single health system might yield results that better generalize to

My focus here is post-revision — where we are after the CR changes? Despite hope that the revised CR would provide a clearer QI-research boundary, the revised CR leaves the same ambiguities largely intact. These ambiguities impose costs on the performance of "high-validity" QI. I shall sketch some guidance for health systems and regulators on operating within the revised Rule, and close with a suggestion for how to modify the Rule. I offer a proposed regulatory exclusion for QI activities, defined as an evaluation performed by a health system to improve the health system's care to its patients, provided such activities impose no risks greater than usual care.

The Journal of Law, Medicine & Ethics, 47 (2019): 238-246. © 2019 The Author(s)

https://doi.org/10.1177/1073110519857279 Published online by Cambridge University Press

other health systems than a QI initiative at a single hospital.

Together, these trends have given QI characteristics associated with research. These characteristics include simple or cluster randomization, multi-system scope, use of outcomes data, and publication of results. QI now "looks" more like research. These characteristics in turn have produced "high-validity" QI. The evolution of QI is primarily one of internal validity, but not entirely. Multi-site QI interventions have produced results that better generalize to other health systems. Further, the evolution continues, as health services researchers partner with their institutions to design evaluations that have the intent of both improving care locally and producing generalizable knowledge.[17]

One framing of these trends is to characterize "high-validity" QI as human subjects research. That is, one view is that QI has become human subjects research, and health systems now conduct research to understand and improve their own care. I caution, however, against abandoning the label of QI because QI is difficult to distinguish from a class of clinical research, one focused on changes to institutional practice. There is still an inward-looking orientation of these activities that remains useful to categorize a set of activities as QI. When a chief quality officer starts an institutional initiative to improve sepsis care for hospitalized patients, implements a series of steps to change practice, and evaluates the impact of those measures, those steps are QI. This practice has different organizational and cultural characteristics from clinical research. The evaluation is conducted in response to an operational priority, the funding sources tend to be operational rather than grant-based, and the intended audience is more likely to be institutional leadership rather than a community of investigators. None of these characteristics in isolation delineates an absolute boundary between research and QI, but they do represent general differences. Yet, while the QI label does describe a class of evaluative activities, it does not address a separate question about the regulatory definition of human subjects research, when that definition may encompass QI, and the implications for conducting high-validity QI.

Common Rule Guidance: Original and Revised

Despite changes in QI practice, the revised Common Rule offers little clarification to help health care systems understand when and how to apply IRB review to QI. The Rule does not apply to QI activities as QI, but the persistent question is when a QI activity is human subjects research under the Rule. Some changes related to such provisions as multi-site IRB review could affect QI.[18] However, the general framework has not changed: the definition of research in the Rule does the work of identifying QI, and that definition is largely unchanged.[19] Since most QI is systematic, QI as research comes to depend on whether the QI activity is "designed to develop or contribute to generalizable knowledge," unless the QI initiative meets one of a limited number of exemptions.[20]

Generalizability, however, is an unclear standard.[21] There are two major sources of ambiguity with generalizability that apply to QI.[22] First, generalizability is a continuum, not a binary, and even were the research definition otherwise clear, it would not be clear when, for example, a randomized, multi-site QI activity becomes sufficiently generalizable to qualify as research. Even a single example from one patient offers lessons for other patients, and even a national, multi-site clinical trial may fail to generalize to specific patient populations or settings of care. QI usually falls in the middle: it incorporates data from many patients, but those patients come from a single institution or set of institutions within a health system. QI's generalizability is contextual, depending on the nature of the intervention under evaluation, patient characteristics, and other factors that vary from one QI initiative to the next, but which determine the degree to which an evaluation is similar to what other institutions might conduct.

Second, the role of intent in generalizability is unclear. The regulatory definition is a systematic investigation "designed to develop or contribute to generalizable knowledge," not one that in fact produces generalizable knowledge.[23] The revised CR retains this phrasing without adding explanation. QI could be distinguished from other classes of high-validity activities by the intent. However, the revised CR leaves unanswered several questions. First, is the intent a subjective one, of the practitioner? Whose subjective intent matters? If the practitioner designs an evaluation only to reach conclusions about local care, but the design is one that will generalize, does it matter what the subjective intent is?

The generalizability provision, while ambiguous, also appears to lack ethical weight in determining the type of review. It is unclear why a QI initiative should incur independent review because of generalizability, especially if generalizing depends on context-dependent factors that seem orthogonal to patient risk. Two important considerations here are the magnitude of risk and whether the subjects themselves, who assume the risks of participating in evaluation, also receive the potential benefits. For the first question, greater generalizability does not increase the magnitude of risk, at the margin. The factors that increase generalizability,

The Journal of Law, Medicine & Ethics, 47 (2019): 238-246. © 2019 The Author(s)

https://doi.org/10.1177/1073110519857279 Published online by Cambridge University Press

such as the representative of the patient population and setting of evaluation to other potential contexts in which the intervention might be delivered, do not in themselves increase risk. More generalizable findings do change the intended beneficiaries — they now enlarge to include other patient populations beyond the local health systems. However, the study environment is still most representative of the local environment. To the extent that study patients may benefit through future use of the local health system, they continue to benefit even if results also inform care for patients of other health systems. Nor do other ethical harms change. If a property of the generalizable design does change the potential harms, then that property, not the generalizability of the design, should influence the sort of review required.

To better illustrate this point, consider two QI scenarios at hospital A that differ only in their degree of generalizability. Both evaluate an early warning system for clinical deterioration placed in the EHR. The warning system uses information about patient vital signs, laboratory values, and other information to predict deterioration in the next six hours. The system's criteria are identical. In both cases, the criteria have never been tested in any systematic fashion. In both cases, hospital A implements the warning system on hospital wards in a staggered fashion to help evaluate its impact on hospital mortality, randomly choosing the order of wards. In case 1, however, hospital A has developed the warning system internally, and no other institution has a similar system. The system differs in multiple ways. It uses a prediction model different from other health systems and it relies on unique organizational resources that are not available at most other health systems. Other institutions use very different warning systems, also developed internally. In case 2, hospital A is using a warning system widely adopted at hospital systems across the country. Case 2 will better contribute to generalizable knowledge. However, case 1 may yield greater risks to patients because there is no evidence, even from informal examples, to help quantify predicted risks. There is thus greater uncertainty regarding the estimated risk to patients of being randomized to the system or control. At a minimum, even if the risks are not greater, they do not appear to be lesser. Nor are there clear additional ethical harms presented by case 2.[24]

Apart from the definition of research, a second important source of ambiguity is how measuring clinical outcomes affects the status of a QI activity under the Rule. By clinical outcomes, I distinguish outcomes from process measures. Process measures refer to changes in care delivery that matter only as intermediate steps to desired changes in health states. Examples include the proportion of clinicians who wash their hands before entering patient rooms, which might be a process measure for an educational campaign to improve handwashing. Clinical outcomes are those that process measures are intended to influence, such as infection rates. Outcomes were not referenced in the original Rule. However, guidance from the Office for Human Research Protections (OHRP) suggested that OHRP interprets the measurement of clinical outcomes to be a relevant factor in determining whether an evaluative activity is research.[25]

The revised CR adds a definition for clinical trials that adds to the indirect evidence that measurement of clinical outcomes may affect an evaluation's regulatory status as research. It defines a clinical trial as a study "in which one or more human subjects are prospectively assigned to one or more interventions (which may include placebo or other control) to evaluate the effects of the interventions on biomedical or behavioral health-related outcomes."[26] I think this definition of a clinical trial is wrong to use measurement of clinical outcomes as a criterion. Measurement of clinical outcomes is neither necessary nor sufficient to qualify an evaluative undertaking as a clinical trial. A randomized study may be designed to test the impact of an intervention on process measures, and process measures may be the primary endpoints in clinical trials. Likewise, an evaluation may measure clinical outcomes in a form that is not generalizable. While this definition has a formal narrow role within the revised Rule for the posting of consent forms, it could contribute to use of clinical outcomes as a factor in classifying a QI activity as a "clinical trial" and thus, by implication, as research regulated by the revised CR. Part of the scope may depend on the interpretation of "assign" and whether it encompasses only specific forms of random assignment, such as parallel patient-level assignment, or whether the scope refers more broadly to systematic methods of assigning patients to evaluation arms.

Even within the regulatory framework focused on generalizability, measurement of clinical outcomes rather than process measures alone should not affect whether a QI activity is designed to generalize. A QI initiative can reach generalizable conclusions about process measures or clinical outcomes. Generalizability about outcomes implies generalizability about process measures. Process measures are intermediate steps that mediate clinical outcomes. If a handwashing educational campaign is intended to reduce infection rates, it does so by increasing handwashing rates. For effects on infection rates to generalize, effects on process outcomes must as well. Conversely, an effect on outcome A works through an effect on some pro-

The Journal of Law, Medicine & Ethics, 47 (2019): 238-246. © 2019 The Author(s)
https://doi.org/10.1177/1073110519857279 Published online by Cambridge University Press

cess measure P. Reaching generalizability conclusions about the effect on P implies similar conclusions about A. Generalizability does not turn on the process versus outcome distinction, and regulatory definitions of generalizable research should not turn on such a distinction either.

Implications of CR Ambiguity and Oversight for QI Activities

The CR criteria for oversight of QI are ambiguous. In many cases, QI activities are reviewed by IRBs.[27] When QI is thought to be research under the CR, QI practitioners submit it for IRB review. Problems with CR oversight are not fatal to the QI enterprise.

Yet, there are costs to CR ambiguity and the unpredictable need for oversight. First, the CR creates the wrong incentives for QI. I stated at the onset that health systems should, subject to considerations about cost, safety, and practicality, pursue high-validity QI. When appropriate, use of high-validity techniques to infer causal impact on important clinical outcomes should be encouraged. At the margin, QI activities are more useful when institutions can reach valid conclusions about the impact of the steps they take to improve. They are more useful still when conclusions speak meaningfully to the needs of other institutions. Institutions pursue QI to improve their own care, but if in so doing, they learn lessons to help other institutions, so much the better. Yet, the need for independent oversight through an institutional review board (IRB) may depend on pursuing high-validity QI rather than low-validity QI.[28]

Uncertainty about IRB requirements imposes costs and misaligns incentives. QI practitioners may be unclear about when they need IRB review, what level of review they need, and how to decide if they should submit for IRB review. In the face of uncertainty, some practitioners choose to ask IRBs to make these determinations. Yet these reviews increase administrative costs to IRBs and cause delays in undertaking QI activities.[29] Practitioners face incentives to "disambiguate" by avoiding high-validity methods, because high-validity relates to the systematic (internal validity) and generalizable (external validity) provisions of the research definition. This uncertainty is not helped by inconsistencies between IRBs.

The IRB process is also ill-adapted to the flexibility needed for QI activities. For example, activities undertaken as QI are more likely than research to use unconventional methods that, while systematic and high-validity, differ in ways that fail to conform to the review demands of an IRB process designed for research. One of these unconventional methods is rapid-cycle evaluation. For example, hospital A designs an early warning system to predict sepsis within six hours. The objective is to improve local recognition and early treatment of septic states. Hospital A decides to conduct a series of short, systematic trials, each followed by iterative modification of the alert and reassessment. These methods pose challenges for IRB review, which is designed for research protocols with a fixed, unchanging design that can be submitted upfront for review. One option is to say that the protocol fails to qualify as research because it is not a systematic investigation, but an iterative process can be systematic and yield generalizable results. It also begs the question: why should the protocol under independent review only if it uses one fixed protocol rather than series of iterative changes?

Third, although I cannot point to systematic evidence, my anecdotal impression is that the CR rules have changed how QI practitioners view their methods. Techniques like randomization, measurement of clinical outcomes, and publication of outcomes are associated with research. These techniques are associated with high internal validity, high external validity QI evaluation. The Rule reinforces this conception through the definition of research, and it is this definition that QI practitioners must use to determine review of their QI activities. QI comes to be distin-

I do not intend here to ignore benefits of independent review. IRB review of QI activities has some benefits. It provides an additional layer of review of risks to patient health, other ethical harms, and ethical implications of protocol design. Data security, for example, is important when health care systems use sensitive information about patients, whether for clinical care or other operational pursuits. My purpose is not to deny any benefits, only to note the costs that accompany the ambiguous relationship that remains with the revised Rule.

The Journal of Law, Medicine & Ethics, 47 (2019): 238-246. © 2019 The Author(s)

https://doi.org/10.1177/1073110519857279 Published online by Cambridge University Press

guished from this conception of research by its reliance on methods that lack these characteristics. As a result, both QI practitioners and outsiders may come to view techniques like randomization as inappropriate because they are not a part of QI, and practitioners are conducting QI, not research. This process is exactly backwards. Instead, QI practitioners should use the techniques appropriate to conduct high-validity QI, subject to operational constraints. When they depart from these methods — and sometimes they should — the rationale should not be that QI simply does not "do" randomization.

All these problems are more significant for the learning health systems. One QI initiative does not incur significant costs in navigating uncertainty and IRB review. A continual process of QI, spread through multiple departments, however, does incur such costs.

I do not intend here to ignore benefits of independent review. IRB review of QI activities has some benefits. It provides an additional layer of review of risks to patient health, other ethical harms, and ethical implications of protocol design. Data security, for example, is important when health care systems use sensitive information about patients, whether for clinical care or other operational pursuits. My purpose is not to deny any benefits, only to note the costs that accompany the ambiguous relationship that remains with the revised Rule. I note also that benefits — along with costs — would accompany independent review of *any* use of patient data, including an independent review of every use of patient data for operational purposes, even when the use does not directly benefit the patient, such as for making decisions about future changes to patient services. Such review would, however, be impractical.

Options in the Current Regulatory Environment

I believe there are steps that health systems and regulators can take to mitigate these costs, while adhering to the Rule and maintaining ethical protections for patients. I turn first to IRBs, and I focus on process, followed by substance. When I refer to IRBs, I mean IRBs working collaboratively with health systems. Learning health systems are the focus of this essay, and their IRBs operate within, or in conjunction with, health systems. IRBs are part of the health system enterprise. Their primary responsibility is to ensure the ethical treatment of research subjects. However, as a part of their obligations to subjects, who are also patients in this context, IRBs have an obligation to work collaboratively with health systems to avoid unnecessary obstacles to high-validity QI, so long as such work does not come into conflict with their other ethical duties to patient. I believe a lot can be done through process to reduce uncertainty, if not eliminate the counter-productive incentives for QI.

First, IRBs should adopt procedural measures to reduce confusion. The first is offering ex ante help to QI practitioners planning QI activities. Some IRBs already take many of these steps.[30] Help can include clear checklists to identify when IRB review needed, as well as specific requirements for level of review and patient consent. Many IRBs have these materials already. Their contents vary from IRB to IRB. I advise for more consistency. IRBs should borrow (with permission) the materials of other IRBs. There are more than 600 health systems in the United States, but there is no need for 600 IRB checklists.[31] We would do better to truncate to a set of ten or fifteen sets of guidelines for different types of health systems. A second step is for IRB staff to work closely with clinical and administrative departments that conduct QI. Such collaboration could help departments to educate their staff on IRB requirements, and how to conduct QI without IRB review becoming an obstacle. QI practitioners, IRB staff, and health system leadership need a common definition of QI to make distinctions, ideally a written definition, one shared by multiple health systems. I offered one in the introduction that would tie QI to the purpose — improving local care — as well as the magnitude of risk (similar or less than usual care). If resources permit, I encourage IRBs to have "QI specialists," IRB staff who review all the applications that arise from QI activities.[32] IRBs can allow applicants to declare an activity to be QI to help with this process.

The review process itself should focus on meeting regulatory requirements while promoting high-validity QI. IRBs already use flexibility to accommodate the needs of specific types of research. Qualitative research, for example, often uses iterative modification of interview guides and other subject materials in response to feedback. Not every edit of materials requires a protocol modified with repeat IRB staff review, and no regulatory requirement mandates review of all changes. Similarly, iterative protocols in QI might not require repeat review. For example, it might be possible for a QI protocol to describe in advance the iterations that will be made in the intervention and how the study design will be modified.

Turning to substance, I suggest that IRBs focus on the definition of research rather than "flags" associated with research, which are tempting to take as per se markers of a research study. Randomization (here used to mean random assignment of interventions) is one example. It is true that many studies regulated as research use random assignment. However, randomization is insufficient to make an activity research. Randomization by itself does not make results of an

The Journal of Law, Medicine & Ethics, 47 (2019): 238-246. © 2019 The Author(s)
https://doi.org/10.1177/1073110519857279 Published online by Cambridge University Press

evaluation more generalizable.[33] Randomization may increase internal validity by removing unmeasured confounders, helping identify the causal connection between an intervention and measured outcomes. However, it does not by itself make that connection more generalizable to other situations. To the extent randomization requires changing routine practice to make randomization feasible, it may even make results less generalizable. A second similar characteristic is measurement of clinical outcomes rather than process measures.

I would encourage OHRP to add guidance to its frequently asked questions on quality improvement, stating that it does not view random assignment and measurement of clinical outcomes as sufficient in themselves to make a systematic investigation research. OHRP previously indicated that publishing results is insufficient.[34] As OHRP notes, people may wish to publish non-research for a variety of reasons. Ironically, an intent to publish offers some degree of evidence that an evaluation is designed to generalize, as something about the evaluation — the methods, results, or purpose — must generalize to make the evaluation useful to other health systems. In contrast, randomization and measurement of clinical outcomes may serve entirely internal purposes. However, there are two reasons that an intent to publish does not necessarily mean an evaluation is research. First, as noted, all forms of evidence offer some degree of generalizability information. Second, a QI evaluation may provide useful lessons for other health systems without necessarily reaching generalizable conclusions. OHRP is thus correct to consider neither an intent to publish nor the actual fact of publication as sufficient to qualify an activity as research. So too could the same be stated of randomization and measuring clinical outcomes. These are common features, like publication, of human subjects research, but they are not definitionally sufficient qualities.

Should Existing Regulations Be Changed?

I suggest further formal changes to the CR to clarify the treatment of quality improvement. There are multiple options to address problems of uncertainty and misaligned incentives. Generally, options fall into one of two approaches, indirect or direct. The indirect approaches would involve clarifying terms such that the boundary between non-reviewable QI and QI as research is easier to identify. This is could involve, for example, adding a formal definition of generalizability. The direct approach would address QI explicitly in the Rule.

I doubt any indirect approach will end ambiguity. I doubt that any definition of generalizability can eliminate uncertainty as applied to QI — there will always be a boundary problem. Generalizability is a concept with imprecise boundaries, and I am not sure that any definition could construct a clear division between QI and research. Nor can it change the incentives. The Rule conceives of research as a high-validity endeavor, one with high internal and external validity. The definition of research as systematic and designed to generalize reflects this understanding.

Instead, I think OHRP should create an explicit exclusion for quality improvement activities. OHRP contemplated such an exclusion in the proposed rule, but discarded the exclusion out of concerns that it would cause additional confusion, and it is not clear why OHRP did not consider a revision definition for exclusion instead. It proposed an exclusion for an evaluation when the purpose was "limited to altering the utilization of the accepted practice and collecting data or biospecimens to evaluate the effects on the utilization of the practice."[35] This definition would have formalized the importance of the distinction between measuring process endpoints and clinical endpoints. For example, a handwashing initiative could be evaluated through a randomized process, but not if the QI activity measured infection rates in addition to handwashing rates. OHRP was right to discard this definition. While it would have reduced uncertainty, it would have done so at the cost of reinforcing incentives for QI to limit measurement to process measures, excluding those clinical outcomes that matter most.

I encourage OHRP to consider a different definition of QI, one based on the underlying objectives. This definition would have two critical components. The first is to connect the doer and the objective to a health system. An evaluation would be excluded when it is performed by a health system primarily to improve the care that the health system provides to its own patients. Second, the definition should limit such activities to those that are minimal risk, but a modified formulation of minimal risk. Risk should be based on what patients experience absent the intervention. The current definition bases risk instead on the magnitude of harms "encountered the performance of routine physical or psychological examinations or tests."[36] For patients with acute or chronic illness, however, the level of risk they encounter in their routine medical care may be quite different. Risk should be based on the anticipated incremental potential harm to an individual based on existing information, and thus to the risk encountered in medical care absent the intervention under evaluation. By the current definition, an intervention could reduce risk for a patient relative to usual care, yet not be considered minimal risk because it still results in a level of risk beyond routine exami-

The Journal of Law, Medicine & Ethics, 47 (2019): 238-246. © 2019 The Author(s)
https://doi.org/10.1177/1073110519857279 Published online by Cambridge University Press

nation. Putting these elements together, I suggest an exclusion for activities conducted by a health care system to improve the care they provide to their patients, when those activities do not increase the frequency or magnitude of harms for patients beyond what they would otherwise experience through clinical care.

This new definition relies on some element of good faith because it bases review on the motivations of practitioners. Could not a researcher claim that a study of a new medical procedure she developed was being conducted only to improve care at the local institution? I believe these concerns are overstated. Health care operations always rely on the good faith of practitioners. Use of patient health information is limited to certain purposes under the HIPAA privacy rule, yet not every use is reviewed by an independent board in advance to ensure it complies with the definition. A degree of trust in the good faith efforts of practitioners is necessary for health care operations to function. Incentives support this level of trust. Most important is to achieve a system of oversight that protects research subjects while permitting learning health systems to use rigorous QI methods to improve the care they provide to their patients.

Note

Dr. Rolnick reports personal fees from Tuple Health Inc. and personal fees from Acumen, LLC, outside of the submitted work.

References

1. Institute of Medicine (US) Roundtable on Evidence-Based Medicine, *The Learning Healthcare System: Workshop Summary*, L. A. Olsen, D. Aisner, and J. M. McGinnis eds., (Washington, DC: National Academies Press (US), 2007), at 6.
2. D. A. Chambers, R. E. Glasgow, and K. C. Stange, "The Dynamic Sustainability Framework: Addressing the Paradox of Sustainment amid Ongoing Change," *Implementation Science* 8, no. 1 (2013): 1-11, at 6..
3. G. Ogrinc et al., "An Instrument to Differentiate between Clinical Research and Quality Improvement," *IRB* 35, no. 5 (2013): 1–8, at 1.
4. M. A. Baily et al., "The Ethics of Using QI Methods to Improve Health Care Quality and Safety," *The Hastings Center Report* 36, no. 4 (2006): S1-40; C. E. Goldstein et al., "Accommodating Quality and Service Improvement Research within Existing Ethical Principles," *Trials* 19, no. 1 (2018): 334-340; N. Kass et al., "Controversy and Quality Improvement: Lingering Questions about Ethics, Oversight, and Patient Safety Research," *Joint Commission Journal on Quality and Patient Safety* 34, no. 6 (2008): 349–353; K. Fiscella et al., "Ethical Oversight in Quality Improvement and Quality Improvement Research: New Approaches to Promote a Learning Health Care System," *BMC Medical Ethics* 16, no. 1 (2015): 63-69.
5. H. H. Pierce and R. E. McKinney, "Opportunities Missed and Created by the New Common Rule," *American Journal of Bioethics* 17, no. 7 (2017): 36–38, at 37.
6. D. A. Chambers, W. G. Feero, and M. J. Khoury, "Convergence of Implementation Science, Precision Medicine, and the Learning Health Care System: A New Model for Biomedical Research," *JAMA* 315, no. 18 (2016): 1941-1942, 1941.
7. J. Henry et al., "Adoption of Electronic Health Record Systems among U.S. Non-Federal Acute Care Hospitals: 2008-2015," *available at* <https://www.healthit.gov/sites/default/files/briefs/2015_hospital_adoption_db_v17.pdf> (last visited Jan. 12, 2019).
8. H. J. Murff et al., "Automated Identification of Postoperative Complications Within an Electronic Medical Record Using Natural Language Processing," *JAMA* 306, no. 8 (2011): 848-855.
9. A. N. Makam, O. K. Nguyen, and A. D. Auerbach, "Diagnostic Accuracy and Effectiveness of Automated Electronic Sepsis Alert Systems: A Systematic Review," *Journal of Hospital Medicine* 10, no. 6 (2015): 396–402.
10. M. R. Cowie et al., "Electronic Health Records to Facilitate Clinical Research," *Clinical Research in Cardiology* 106, no. 1 (2017): 1–9.
11. J. Rolnick et al., "Ethical Oversight of Quality Improvement and the Research-QI Boundary: A New Common Rule Changes Little," *IRB: Ethics & Human Research* 39, no. 3 (2017): 1–10; C. A. Longhurst, R. A. Harrington, and N. H. Shah, "A 'green Button' for Using Aggregate Patient Data at the Point of Care," *Health Affairs (Project Hope)* 33, no. 7 (2014): 1229–1235, at 1232.
12. J.A. Hirsch et al., "PQRS and the MACRA: Value-Based Payments Have Moved from Concept to Reality," *American Journal of Neuroradiology* 37, no. 6 (2016): 2195-2200. Inpatient acute care is a partial exception to fee-for-service. Care is bundled under the DRG system. However, even there, hospitals have flexibility to use different DRGS of varying reimbursement and intensity of services.
13. D. M. Cutler, "Rising Medical Costs Mean More Rough Times Ahead," *JAMA* 318, no. 6 (2017): 508-509, at 509.
14. "The Impact of Hospital Consolidation on Medical Costs," NCCI Holdings Inc., *available at* <https://www.ncci.com/Articles/Pages/II_Insights_QEB_Impact-of-Hospital-Consolidation-on-Medical-Costs.aspx> (last visited Jan. 12, 2019).
15. H. B. Fleishon et al., "Academic Medical Centers and Community Hospitals Integration: Trends and Strategies," *Journal of the American College of Radiology* 14, no. 1 (2017): 45–51.
16. T. Xu, A. W. Wu, and M. A. Makary, "The Potential Hazards of Hospital Consolidation: Implications for Quality, Access, and Price," *JAMA* 314, no. 13 (2015): 1337-1338.
17. A. Hofmeyer, C. Scott, and L. Lagendyk, "Researcher-Decision-Maker Partnerships in Health Services Research: Practical Challenges, Guiding Principles," *BMC Health Services Research* 12, no. 1 (2012): 280-286.
18. 45 C.F.R. 46.114 (2018).
19. This same definition is used for other federal regulations that affect QI activities, including the definition in Health Insurance Portability and Accountability Act of 1996. 45 CFR 164.501 (2018). The definition of research is unchanged, but the revised CR definition includes a set of activities that are not research. 45 C.F.R. 46.102(l) (2018).
20. *Id.* The most important is an exemption for secondary analysis of identifiable health information regulated by the HIPAA privacy rule. 45 C.F.R. 46.104(d)(4)(iii) (2018). This category of exemption does not require a limited IRB review. However, certain requirements, such as informed consent, may still apply under the privacy rule, and involve a similar inquiry into the application of the definition of research.
21. See Rolnick et al., *supra* note 11; Pierce and McKinney, "Opportunities Missed and Created by the New Common Rule," 17, no. 7 (2017): 36.38; See Baily et al., *supra* note 5, at 37.
22. Applying the generalizability standard is not a unique problem for QI, but QI is an important example of the problems.
23. 45 C.F.R. 46.102(l) (2018).
24. Case 1 may involve evaluation of an untested intervention, and therefore not QI. I would dispute that a historical record of testing is what distinguishes QI from research. Many changes to systems practice will be untested, and their contextual nature means that any local system change is, in a sense, untested. Regardless, the point here is to illustrate that

The Journal of Law, Medicine & Ethics, 47 (2019): 238-246. © 2019 The Author(s)

https://doi.org/10.1177/1073110519857279 Published online by Cambridge University Press

generalizability does not by itself lead to ethical harms that would justify independent review.

25. Quality Improvement Activities FAQs, United States Department of Health and Human Services, Office for Human Research Protections, *available at* <http://www.hhs.gov/ohrp/policy/faq/quality-improvement-activities/index.html> (last visited Jan. 1, 2019). It is unclear how OHRP is distinguishing the quality of care from patient outcomes, or why measurement of patient outcomes rather than other endpoints (e.g. a process measure) should determine whether an evaluation of an untested intervention is research under the Rule.
26. 45 C.F.R. 46.102(b) (2018).
27. D. Whicher et al., "The Views of Quality Improvement Professionals and Comparative Effectiveness Researchers on Ethics, IRBs, and Oversight," *Journal of Empirical Research on Human Research Ethics* 10, no. 2 (2015): 132–144, at 138.
28. The Common Rule only applies to "all research involving human subjects conducted, supported, or otherwise subject to regulation by any Federal department or agency." 45 C.F.R. 46.101(a) (2018). However, as a practical matter, funded institutions are likely to apply it to any research activities, regardless of direct funding source.
29. Infectious Diseases Society of America, "Grinding to a Halt: The Effects of the Increasing Regulatory Burden on Research and Quality Improvement Efforts," *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* 49, no. 3 (2009): 328–335.
30. For example, see "Quality Improvement Activities in Health Care Versus Research," Duke University Health System, Human Research Protection Program (April 30, 2013), *available at* <https://irb.duhs.duke.edu/sites/irb.duhs.duke.edu/files/QI_policy_and_checklist_4-30-2013jw.pdf> (last visited Jan. 1, 2019).
31. Snapshot of U.S. Health Systems, 2016, Agency for Healthcare Research and Quality, *available at* <https://www.ahrq.gov/sites/default/files/wysiwyg/snapshot-of-us-health-systems-2016v2.pdf> (last visited Jan. 1, 2019).
32. It may be useful to have a QI "checkbox" on applications.
33. See Rolnick, *supra* note 11, at 7.
34. Quality Improvement Activities FAQs, United States Department of Health and Human Services, Office for Human Research Protections, *available at* <http://www.hhs.gov/ohrp/policy/faq/quality-improvement-activities/index.html> (last visited Jan. 1, 2019).
35. Federal Policy for the Protection of Human Subjects, 80 Fed. Reg. 53933, 53948 (Sept. 8, 2015).
36. 45 C.F.R. 46.102(j) (2018).

The Journal of Law, Medicine & Ethics, 47 (2019): 238-246. © 2019 The Author(s)

https://doi.org/10.1177/1073110519857279 Published online by Cambridge University Press