

Measurement Error in Performance Studies of Health Information Technology: Lessons from the Management Literature

A.S. Litwin¹; A.C. Avgar²; P.J. Pronovost³

¹Carey Business School and School of Medicine, Johns Hopkins University, Baltimore, MD; ²School of Labor & Employment Relations and College of Medicine, University of Illinois at Urbana-Champaign, Urbana-Champaign, IL; ³School of Medicine, Bloomberg School of Public Health, and Carey Business School, Johns Hopkins University, Baltimore, MD

Keywords

Health information technology, electronic health records, research methods, organizational behavior

Summary

Just as researchers and clinicians struggle to pin down the benefits attendant to health information technology (IT), management scholars have long labored to identify the performance effects arising from new technologies and from other organizational innovations, namely the reorganization of work and the devolution of decision-making authority. This paper applies lessons from that literature to theorize the likely sources of measurement error that yield the weak statistical relationship between measures of health IT and various performance outcomes. In so doing, it complements the evaluation literature's more conceptual examination of health IT's limited performance impact. The paper focuses on seven issues, in particular, that likely bias downward the estimated performance effects of health IT. They are 1.) negative self-selection, 2.) omitted or unobserved variables, 3.) mismeasured contextual variables, 4.) mismeasured health IT variables, 5.) lack of attention to the specific stage of the adoption-to-use continuum being examined, 6.) too short of a time horizon, and 7.) inappropriate units-of-analysis. The authors offer ways to counter these challenges. Looking forward more broadly, they suggest that researchers take an organizationally-grounded approach that privileges internal validity over generalizability. This focus on statistical and empirical issues in health IT-performance studies should be complemented by a focus on theoretical issues, in particular, the ways that health IT creates value and apportions it to various stakeholders.

Correspondence to:

Adam Seth Litwin, PhD, MSc
Johns Hopkins University
100 International Drive
Baltimore, MD 21202-1099
email: aslitwin@jhu.edu
phone: (410) 234-9420
fax: (410) 234-9439.

Appl Clin Inf 2012; 3: 210-220

doi:10.4338/ACI-2012-02-R-0004

received: February 16, 2012

accepted: May 20, 2012

published: June 13, 2012

Citation: Litwin AS, Avgar AC, Pronovost PJ. Measurement error in performance studies of health information technology: Lessons from the management literature. *Appl Clin Inf* 2012; 3: 210-220
<http://dx.doi.org/10.4338/ACI-2012-02-R-0004>

1. Introduction

In a sea of controversy and disagreement over how to reform the US healthcare system, the one island of apparent consensus has been health information technology (IT) – namely, that this technology will play a key role in improving quality, safety, and efficiency in care delivery. Policymakers, above all, have accepted this notion, allocating \$14-\$27 billion in economic stimulus funds to encourage investments in health IT, before even attempting to take on full-fledged reforms [1]. And, nongovernmental actors have also begun undertaking sizable investments of their own in health IT [2]. This is curious, since those few early performance studies that looked beyond a small number of well-resourced or structurally unique institutions [3, 4], where they even detected positive performance effects, generally found the performance effects of health IT to be clinically, if not statistically, insignificant [5–8].

There are a number of complementary paths toward explaining this disconnect between health IT investments and real outcomes. On the one hand, the problem requires deeper consideration of the theories linking health IT to outcomes, in particular, the logic by which the benefits of health IT are generated, partitioned, and appropriated by various stakeholders [9]. This approach is itself served by a well-developed and far-reaching literature focused on improving the evaluation of health IT, largely in clinical settings [10–13]. In most cases, these analyses result in a set of prescriptions for study design and administration, highlighting especially common challenges that researchers must confront, too often in the post-deployment stage of evaluation. However, despite their wide scope, these studies stop short of explaining in precise, statistical terms how any particular challenge outlined in their models would mechanically impact a quantitative estimate of health IT's effectiveness. Furthermore, all of these studies call for research approaches from other disciplines to aid in the advance and application of their frameworks, with a particular call for methodological support and guidance.

This article aims to transcend the conceptual approach taken by these studies and accept the challenge they pose to interdisciplinarians. Specifically, we consult the management literature to examine the statistical implications of many of the conceptual issues raised in the evaluation literature. In so doing, we explore sources of measurement error that inhere in those health IT-performance studies that admirably seek to transcend the generalizability bounds attendant to case studies, but that ultimately find effects to be inconclusive or weak.

Indeed, the phenomenon of new technology's "missing" performance effects is one that management scholars have had to confront on a number of occasions. At least two crisp examples come to mind. The first involves the US automobile industry's early investment in automation and its apparent association with productivity declines. General Motors, in particular, invested \$650 million in a single factory without realizing performance improvements [14]. Analyses revealed that GM's blatant but shrewd tactic to replicate Japanese sources of productivity and performance missed a key element. It was actually the combination of new technologies and innovative employment practices that positioned shop floor workers to "give wisdom to the machine" [15, 16], a potent complementarity that was shown to deliver results not only in other areas of manufacturing [17] and in the service sector [18], but even for nonclinical functions in a healthcare setting [19]. The second example, which has since been labeled the "productivity paradox" [20], is often encapsulated by Nobel laureate Robert Solow's assertion that "[one] can see the computer age everywhere but in the productivity statistics" [21]. More specifically, those studying IT investment struggled to explain why businesses directed enormous amounts of capital into hardware and software despite scant financial returns [22]. Likewise, those focused on employee involvement and other innovative employment practices found it difficult to identify the performance effects arising from these "high-performance" work practices [23, 24]. But, if team-based production or other employment practices that transfer authority to workers cannot be linked to improved productivity, then why did managers continue to implement them? Ultimately, these anomalies were woven together to fabricate a cohesive explanation for both of them. In short, it turned out that IT investments *in conjunction with* organizational changes brought about sizable productivity gains, but empirical analyses were not examining IT variables and organizational variables *in concert* nor were studies focused on the appropriate unit-of-analysis – the very organizations in which these production inputs are brought together. Empirical analyses taking this wider and deeper view have succeeded at identifying the performance returns

to both production “inputs.”

This article draws from the management literature in an effort to dispose of the health IT-performance paradox. We first outline some of the statistical obstacles confronting performance studies of health IT, with reference to both health IT evaluation studies and management research that touches upon them. Then, we make explicit recommendations for moving forward.

2. Archetypal Research Design in Health IT-Performance Analyses

The archetypal research design for identifying the performance effects of new technologies or of innovative employment practices is one in which some measure of performance is estimated as a function of a number of variables, one of which is the focal technology or organizational measure. In the case of a health IT-performance study, the researcher has a large, cross-sectional dataset of “units” – be they hospitals, wards, individual providers, medical offices, etc. – some with the technology and some without it. In order to “tease out” the performance effects of health IT, the researcher would employ an ordinary least squares (OLS) technique or one closely related to it that allows for the control of other variables besides IT that influence the outcome of interest. The researcher would be estimating an equation along the lines of

$$y_i = \alpha + \beta_{\text{HIT}} \text{HIT}_i + \beta_{\text{controls}} \text{CONTROLS}_i + \varepsilon_i \quad ,$$

where for each observed unit, i , y , is the outcome of interest, HIT is a binary variable equal to one for those facilities that have the focal technology, CONTROLS is a vector of other factors that potentially drive y besides HIT , and ε is a zero-expectation residual. The parameter of interest in this case is β_{HIT} . Assuming that y represents a “normal” outcome such as clinical care quality (as opposed to an “inferior” outcome, such as cost, in which positive movements are undesirable), then a finding of $\beta_{\text{HIT}} > 0$ implies that there is a mean difference in performance between those facilities with health IT and those without it, net of controls, such that those with health IT perform better than those without it.

The approach just described represents a simplified version of the approach that many of the best health IT-performance studies to date have taken [5, 25], yet most have yielded an estimate of β_{HIT} that is trifling in magnitude, insignificantly different from zero, or present for only a subset of the outcomes examined. So, assuming that researchers and policymakers are right about the performance prospects for health IT – just as managers were right about the benefits of IT and of innovative employment practices, the key question is this: if in fact, $\beta_{\text{HIT}} > 0$, what could engender an estimate of $\beta_{\text{HIT}} \leq 0$?

3. Sources of Downward Bias in Health IT-Performance Analyses

3.1 Negative Self-Selection

There are a number of ways for an estimated health IT performance effect to be biased downward ($\hat{\beta}_{\text{HIT}} \leq \beta_{\text{HIT}}$). Those we discuss in this paper are highlighted in ► Figure 1. First among these is the fact that the health IT “treatment” is not randomly assigned or allocated across the sample [26]. This increases the likelihood of negative self-selection, a genuine threat where true random assignment is simply not possible. That is, $\hat{\beta}_{\text{HIT}}$ will be biased downward if poorly-performing or relatively unsuccessful facilities are more likely than others to invest in health IT [27]. There are at least two mechanisms that could generate this selection issue. On the one hand, it could be that poorly-performing facilities view health IT as a means for “digging themselves out of a hole” [28]. Alternatively, it could be that the most successful facilities are reluctant to make material changes in how the organization delivers care, believing that its incumbent systems and technologies are responsible for its success.

3.2 Unobserved Variables

In the case of negative selection, unobserved or unmeasured variables impinge on both the health IT adoption decision and the subsequent association between health IT adoption and the dependent variable measuring performance [10, 13]. Therefore, it is just a special case of a broader set of forces potentially biasing health IT-performance estimates, the issue of omitted variables. Of course, there are always omitted variables, particularly outside of lab experiments, so their mere existence cannot be allowed to negate all findings. However, there are two situations in which the presence of omitted variables can bias β_{HIT} toward zero. First, it could be that the omitted variable is positively correlated with y and negatively correlated with HIT . For example, a given hospital or clinic may serve a patient base with a static, routine, and well-met set of needs. Second, the downward bias could be the result of an omitted variable that is negatively correlated with y and positively correlated with HIT [29]. For example, a highly-centralized, “command-and-control” management style – the opposite of innovative employment practices – may increase the likelihood of investment while depressing performance. More simply, consider the workflow disruption associated with health IT implementation that, while possibly unobserved, will be positively associated with the incidence of health IT and negatively correlated with performance [30].

3.3 Mismeasured Co-Variates

A primary reason that researchers omit certain variables from their analysis is that they are difficult to measure, particularly for large- n studies in which the researcher has not personally observed the behavior generating the data or settings that involve variables, such as context variables, for which the field may lack validated measurement tools. One can imagine this challenge to generalizability [31] being an especially acute issue in survey research. What makes it especially problematic is that even attempts to measure what had previously been regarded as “unobservable,” such as organizational features or changes to workflows, do not fully remove the bias from β_{HIT} [32]. If, for example, the incidence of something relatively objective and apparent such as health IT is measured with less error than the previously “unobserved” organizational variables, then measurement error associated with the latter creates a situation similar to its not having been measured at all. That is, the estimate for HIT can be biased in the same way as if the hard-to-measure variables were still simply omitted from the analysis altogether.

3.4 Mismeasured Health IT Variables

Even if omitted or mismeasured co-variates are not working to bias downward the effects of health IT, mismeasurement of the IT variable itself could attenuate the estimated effects of health IT [32]. As Poon et al. note [33], many studies tend to measure the incidence of health IT as a binary variable – each observed unit either has the technology in place or does not [34, 35], or with only slightly more nuance by measuring health IT trichotomously rather than dichotomously [5]. This effectively ignores variation related to the extent of actual diffusion or use of the technology within the organization [10, 26, 36]. It also shines light on the importance of examining specific health IT functions or capabilities [31]. That is, it may not be careful enough to study computerized physician order entry (CPOE) or e-prescribing *per se*. Rather, what exact sorts of orders or prescriptions once submitted on paper are now processed electronically? Finally, the measure may fail to account for the degree of decision support or the extent to which the IT system is used for organizational learning, the factors believed to provide a large proportion of the benefits from IT systems.

3.5 Stage of Adoption-Use Continuum

Along the same lines, studies must be sensitive to the specific activity they are measuring with regard to health IT. That is, are they measuring adoption, implementation, or use? Whereas mere adoption will not necessarily improve quality, it is more conceivable that actual use would [10, 36]. One can imagine that the health IT measure in each stage links to a different set of appropriate performance outcomes, and that these outcomes will only start to vary as hypothesized after a certain amount of

time passes. In fact, β_{HIT} could be biased downward if one tries to link adoption or investment directly to performance without first considering the impact of the technology on use. This could occur either through a process of negative self-selection or a case of mismeasured independent variables.

3.6 Time Horizon

Temporal aspects of the model linking health IT to outcomes can also undermine attempts to measure performance. Investments in new technology – what many may see as a key measure of “adoption” generally occur “upfront” and are easily and precisely measured. On the other hand, there are delays in key indicators on both the left- and right-hand sides of the performance equation. That is, organizations are generally slower to make necessary organizational changes around new technologies [22, 32], and these lags necessitate that health IT be given more time to reveal itself in performance indicators [11]. That is, the benefits of IT systems are likely to emerge over time as the organization develops the capacity to fully use the technology [29], so studies with shorter time horizons are likely to yield weaker results [22, 32, 37].

3.7 Unit-of-Analysis

Closely related to the specific activity being studied is the appropriate unit-of-analysis – the facility, department, or perhaps even the individual provider or patient. It depends critically on what stage of the process is being investigated. For example, adoption (or investment) probably occurs at the organizational or firm level as opposed to at the facility level. However, as suggested above, the extent to which health IT gets used likely hinges on variation at a lower level of analysis such as the department, the ward, or even the individual user [10]. If one were to measure health IT as being “on” in a given hospital, then some outcomes measured at the ward level may be wrongly assigned with respect to the health IT variable [28], again biasing β_{HIT} toward zero. This sort of an error can occur quite easily, since organizational-level measures are more likely to be available than sub-organizational ones. Organizational-level measures are also more likely to be measured in dollars than are sub-unit measures, making them much easier to compare to other units in the sample and of broader appeal to policymakers.

4. Addressing Downward Bias in Health IT-Performance Analyses

4.1 Negative Self-Selection and Variables That Are Unobserved or Poorly-Measured

Since negative self-selection is caused by omitted variables and the effect of omitting a variable is akin to the impact of mismeasuring it, the resulting downward pressure on β_{HIT} can be countered similarly irrespective of which of these factors is driving it. From a statistical standpoint, the first step should be to consider the use of panel data – multiple observations of the same unit over time. With these data in hand, one can estimate an equation similar to the one presented above, except that each unit can essentially be used to control for all of the time-constant, unobserved variables that could have been biasing β_{HIT} in the cross-sectional estimates, i.e., a “unit fixed-effects” model. That is, if the omitted variables are relatively stable over the time period studied, then one can use longitudinal data to examine whether *changes* in health IT status appear to generate movements in performance. This method could be a reasonable one for dealing with unobserved or difficult-to-measure aspects of managerial skill, worker quality, or even characteristics of the patient population, and has been employed as means of dealing with issues of selection and of omitted variables [6, 35]. Fixed-effects or panel models are also valuable from a diagnostic perspective. If the β_{HIT} that emerges from a fixed-effects model is of roughly the same magnitude as the β_{HIT} from the conventional, cross-sectional model, then one can infer that time-constant unobservable ascriptions of the units-of-analysis are not the source of health IT’s “missing” performance benefits.

Unfortunately, panel data are no panacea. First, even a relatively small amount of measurement error in the health IT measure, for example, can further bias β_{HIT} toward zero [38]. Second, to the extent there is any positive self-selection, in which high-performing units are especially likely to adopt health IT, adopting organizations, having less room to climb the performance curve, may find it especially difficult to improve their performance. Third, the tradeoff for purging the model of unmeasured variation in time-constant unobservables is the inability to estimate the impact of any of the time-fixed covariates on the dependent variable. For example, if a given hospital provides data on square footage only once or remains in the same region over the period studied – both very reasonable assumptions – then fixed-effects estimates using panel data cannot reveal anything about the relationship between size and y or between region of operation and y . Finally, even if the performance effects of these co-variables were of secondary concern to the researcher, it is likely that the “choice” to rely on cross-sectional rather than panel data was dictated by resource constraints and practicality.

When panel data are not an option, the best way to treat issues of negative selection and of omitted variables more broadly is to measure and include as many of these variables as possible in the vector of controls [10]. Regarding negative self-selection, in particular, researchers cannot consider the link between health IT and performance without a detailed understanding of the drivers of health IT adoption or diffusion. That is, they must draw upon the literature highlighting the organizational differences between adopters and non-adopters, including, for example, the level and sources of practice or hospital revenue [39]. In order to tackle selection bias in performance studies, researchers must identify links between past performance and health IT adoption as well as some of the other factors that appear to drive health IT adoption, since many of these same factors likely influence performance as well [27, 40]. Moreover, studies focused solely on diffusion can offer insights into variables that ought to be modeled in performance studies, even if they are not sources of selection *per se*. Consider, for example, a detailed case study of those facilities that have adopted and then abandoned health IT. This could reveal the ways that health IT is a poor fit with some key characteristic of the facility or the organization that operates it, or it could reveal the nuances of the implementation process that drove its failure.

As noted earlier, perhaps the most overlooked variables in health IT-performance studies are those that were previously overlooked by management researchers seeking to link hardware and software to economic performance. These are variables that can be described as organizational or work-related [19], the omission of which can bias β_{HIT} downward, even negative in the case of Han et al.’s study [41] on CPOE. As noted earlier, management researchers eventually clarified that IT in and of itself is generally not the source of measurable performance gains [19, 32]. Rather, gains arise from complementarities between innovative work practices, including the reorganization of work, and the technology itself. The most direct way to integrate these measures into performance studies would be to add questions about training, employee involvement structures, and workflow redesign to interview protocols and surveys. Unfortunately, these sorts of variables are often measured with a relatively high degree of subjectivity and uncertainty [42], a problem exacerbated by the fact that the respondents with reliable information on health IT adoption and performance may not have a bird’s eye view of structures and processes defining the context in which the IT is used. Therefore, the best solution may be to sacrifice generalizability and sample size by constraining the sample to one in which these contextual measures can be assumed constant. Indeed, studies that have taken this approach have found hypothesized performance effects in samples constrained to 15 outpatient clinics in the same greater metropolitan area [19] or in 41 hospitals located in the same state [43], for example.

This approach also allows for more nuance in measuring the health IT variable. First, even those components of health IT that are well-understood differ in important ways. Therefore, it is not enough to identify the specific type of health IT under study, e.g., CPOE, e-prescribing. Not all systems for CPOE, for example, are created equal in terms of interface design or usability [30]. Second, no matter what form or functionality of health IT is under study, no facility makes a single, discontinuous jump from not having the technology to having it fully implemented and optimally deployed. Precisely how to measure the extent, degree, or level of health IT across a sample will depend on the nature of the focal health IT functionality, the types of outcomes it is expected to influence, and the nature of the mathematical function linking the technology to the particular outcome (e.g., linear, curvilinear).

4.2 The Stage, the Time Horizon, and the Unit

In addition to a detailed understanding of the specific functionality under study, researchers should pay close attention to which stage or stages of the adoption-use continuum they are capturing with their data and their design. As alluded to above, mere adoption, measured by investment, for example, does not have a tightly-defined theoretical relationship to performance measures like care quality, as actual use assuredly moderates this relationship [33, 36]. Therefore, in order to establish $\beta_{\text{HIT}} > 0$, studies should first work to link adoption to use, and then actual use to performance. Furthermore, intra-stage sources of moderation or mediation should be identified, measured, and modeled. Among those most overlooked is time. That is, even after controlling for important contextual variables for a given stage, there is still likely to be a delay between when the health IT is invested in, turned on, or otherwise activated, and when it starts to predictably drive variation in the dependent variable [22, 35]. Therefore, studies must allow enough time for the organization to make necessary changes and to undergo necessary learning around new technologies before taking a “post-” health IT performance measure.

The appropriate unit-of-analysis also depends in great part on the stage being studied. However, if the goal is to establish a reliable and unbiased estimate of the effects of health IT, then the researcher should seek to disaggregate the data to the lowest possible level. This allows for the control of contextual variables at the lowest possible level as well as at each of the levels above it. In other words, if a performance measure is available at the level of the individual patient, for example, then the researcher could control for the patients’ co-morbidities, physicians’ job tenure, characteristics of the ward, and characteristics of the hospital and of the nursing staff that treated the patient, and the health IT measure could be allowed to vary at any or all of these levels. Such an analysis calls for a more sophisticated approach than the conventional cross-sectional estimation used as the basis for the discussion here. Rather, it calls for a multilevel estimation technique, along the lines of hierarchical linear modeling (HLM) [44], which while ideal for advancing empirical analyses of health IT’s performance benefits, is beyond the scope of this paper.

In those situations where the unit-of-observation is too high to account for within-unit variation, one way to mitigate the forces biasing β_{HIT} is to rely on an especially well-informed respondent [45]. Ideally, this respondent will be expert or at least well-positioned to provide information on the technology, performance measures, or some other measure. However, if one person is asked to provide all of the data on “inputs” and “outputs,” there is the possibility that he or she will provide erroneous information on at least one of them. Therefore, where possible, researchers should tap multiple respondents with potentially different functional or hierarchical vantage points for providing information on health IT as well as on contextual variables [46].

5. Health IT-Performance Analyses Moving Forward

Received health IT-performance studies have yielded less-than-conclusive findings on the effectiveness of health IT, much as management research once did in its attempts to link new technologies and new organizational practices to performance measures. This enumeration of common sources of measurement error, made possible through rearview consideration of the management literature, and the subsequent indirect critique of existing health IT-performance studies is not meant to denigrate received empirical work, but to guide those looking to improve upon it. While it is likely that no single study can adhere to all of the lessons gleaned from the analysis above, they can collectively “chip away” at these issues to further clarify the instrumentality of health IT over various outcomes, both clinical and administrative. Among the most important takeaways is the need to consider drivers of diffusion when modeling performance, as there are almost assuredly issues of selection in performance studies that do not. Second, there is an immediate need for research designs that allow the researchers to personally observe the units under study wherever possible as a means of picking up more variables that are typically omitted or mismeasured. In broad terms, we are at a stage of research and policymaking in which the establishment of internal validity must trump efforts toward generalizability. We have not yet shown that these technologies work. So, studies should be focused and grounded, with deep understanding of the structures and processes generating the data, imply-

ing that a prime space in this research stream should still be reserved for rich, qualitative case studies and deep, mixed-methods analyses that ground cleanly-measured organizational and health IT variables in well-understood contexts [47].

This analysis certainly has its own limits. First, it is framed around the weak results typical of earlier studies on health IT. Indeed, a comprehensive review of the most recent studies shows “predominantly positive results” [48]. However, the studies included in the review should also be read with an eye toward the points raised above, many of which can be inverted to illuminate sources of positive bias (e.g., erroneously revealing a generalizable, positive effect) in health IT studies. Most critical would be a tendency toward positive self-selection, in which the most well-resourced organizations with the best performance prospects are also the most likely to have invested in new technology. Similar analogs can be made to the other sources of downward bias raised above. A second challenge to this article is that it focuses exclusively on empirical issues rather than theoretical ones. It is true that for policymakers, knowing the effect of health IT on various productivity and cost outcomes might suffice. But, for hospital administrators, practice managers, and others in individual organizations, financial performance is what really determines success. And, given the ways that value is generated from health IT and then partitioned amongst stakeholders, even technologies that boost real, objective productivity measures may have little, no, or even negative effects on financial performance. That is, even if the net gain to society is positive, there will be winners and losers, and it could be that the losers have a lot of power to prevent systemic, institutional, and even organizational change. These issues have consequences for the health IT-performance link, and are at least as important as the empirical ones taken up here.

Acknowledgements

The authors acknowledge the valuable research assistance of Trevor DaRin and William Graves as well as comments and suggestions from Joan Ash, PhD, MLS, MS, MBA, Harold Lehmann, MD, PhD, and seminar participants at the 2011 Academy of Management annual meetings in San Antonio, TX.

Clinical Relevance Statement

Consumers and providers of care are eager to pin down statistically the relationship between health IT and performance outcomes, including clinical ones. This review draws from the management literature to enumerate the seven most common sources of measurement error in such studies. It offers concrete examples of how these biases come about and how they can be mitigated.

Issue Identified	Example	Suggestion for Mitigation
negative self-selection	§ Poorly-performing facilities view health IT as a way of "digging themselves out of a hole." § Successful facilities are reluctant to overhaul systems and technologies.	
omitted or unobserved variables	§ Hospital or clinic serves a static patient base. § Hospital or clinic employs a "command and control" management style. § Workflow disruptions are associated with health IT adoption.	§ Observe each unit multiple times as a means of controlling for all time-constant, unobserved variables. § Include as many theoretically- or case-justified control variables as possible, with guidance from qualitative data-gathering and time spent "on the ground."
mismeasured contextual variables	§ Organizational factors are measured less reliably than technology incidence.	§ Constrain the scope of the study, sacrificing generalizability for improved causality.
mismeasured health IT variables	§ Incidence of technology measured discretely rather than continuously. § Health IT systems cast as monolithic rather than as an amalgam of well-defined functional modules.	
lack of attention to stage of adoption	§ IT adoption is linked directly to quality outcomes, irrespective of actual use.	§ Link adoption to use, use to quality, and quality to financial outcomes.
short time horizon	§ Observation is done too early, so that the focal health IT system is in place, but without the requisite organizational changes.	§ Be sure that variation in the independent and dependent variables are each measured at the appropriate level, e.g., investment at a higher level than actual use. § Allow time for organizational changes, e.g., workflow redesign, organizational learning, to take place before settling on a "post-health IT" observation of outcomes. § Rely on multiple, "informed" respondents to ensure that independent and dependent variables are all measured reliably rather than conveniently.
inappropriate units-of-analysis	§ Health IT use is measured at the facility level, though actual use within the facility varies greatly across wards	

Fig. 1 Sources of downward bias in health IT performance studies and specific suggestions for mitigation

References

1. Blumenthal D. Launching HITECH. *N Engl J Med* 2010; 362(5): 382–385.
2. Jha AK, DesRoches CM, Kralovec PD, Joshi MS. A progress report on electronic health records in US hospitals. *Health Aff* 2010; 29(10): 1951–1957.
3. Chaudhry B, et al. Systematic review: Impact of health information technology on quality, efficiency, and costs of medical care. *Ann Intern Med* 2006; 144(10): 742–752.
4. Garrido T, Jamieson L, Zhou Y, Wiesenthal A, Liang L. Effect of electronic health records in ambulatory care: Retrospective, serial, cross sectional study. *Br Med J* 2005; 330(581).
5. DesRoches CM, et al. Electronic health records' limited successes suggest more targeted uses. *Health Aff* 2010; 29(4): 639–646.
6. Parente ST, McCullough JS. Health information technology and patient safety: Evidence from panel data. *Health Aff* 2009; 28(2): 357–360.
7. Himmelstein DU, Wright A, Woolhandler S. Hospital computing and the costs and quality of care: A national study. *Am J Med* 2010; 123(1): 40–46.
8. Koppel R, et al. Role of computerized physician order entry systems in facilitating medication errors. *J Am Med Assoc* 2005; 293(10): 1197–1203.
9. Avgar AC, Litwin AS, Pronovost PJ. Using management research to conceptualize the drivers and barriers to health IT adoption. Working paper. 2012.
10. Ancker JS, Kern LM, Abramson E, Kaushal R. The triangle model for evaluating the effect of health information technology on healthcare quality and safety. *J Am Med Inform Assn* 2012; 19(1): 61–65.
11. Ammenwerth E, et al. Visions and strategies to improve evaluation of health information systems: Reflections and lessons based on the HIS-EVAL workshop in innsbruck. *Int J Med Inf* 2004; 73(6): 479–491.
12. Nykänen P, et al. Introducing guidelines for good evaluation practice in health informatics. In: Adlassnig K, Blobel B, Mantas J, Masic I, editors. *Medical Informatics in a United and Healthy Europe*. Amsterdam; Washington, DC: IOS; 2009: 958–962.
13. Westbrook JL, et al. Multimethod evaluation of information and communication technologies in health in the context of wicked problems and sociotechnical theory. *J Am Med Inform Assn* 2007; 14(6): 746–755.
14. Kochan TA. On the human side of technology. *ICL Tech J* 1988; 6(2): 391–400.
15. MacDuffie JP, Krafcik JF. Integrating technology and human resources for high-performance manufacturing: Evidence from the international auto industry. In: Kochan TA, Useem M, editors. *Transforming Organizations*. New York: Oxford; 1992: 209–226.
16. MacDuffie JP. Human-resource bundles and manufacturing performance: Organizational logic and flexible production systems in the world auto industry. *Ind Labor Relat Rev* 1995; 48(2): 197–221.
17. Kelley MR. Participative bureaucracy and productivity in the machined products sector. *Ind Relat* 1996; 35(3): 374–399.
18. Batt RL. Work organization, technology, and performance in customer service and sales. *Ind Labor Relat Rev* 1999; 52(4): 539–564.
19. Litwin AS. Technological change at work: The impact of employee involvement on the effectiveness of health information technology. *Ind Labor Relat Rev* 2011; 64(5): 863–888.
20. Brynjolfsson E, Hitt LM. Beyond computation: Information technology, organizational transformation and business performance. *J Econ Perspect* 2000; 14(4): 23–48.
21. Solow R. We'd better watch out. *New York Times Book Review* 1987 July 12 (36).
22. Brynjolfsson E, Hitt LM. Computing productivity: Firm-level evidence. *Rev Econ Stat* 2003; 85(4): 793–808.
23. Cappelli P, Neumark D. Do 'High-performance' work practices improve establishment-level outcomes? *Ind Labor Relat Rev* 2001; 54(4): 737–775.
24. Freeman RB, Kleiner MM. Who benefits most from employee involvement: Firms or workers? *Am Econ Rev* 2000; 90(2): 219–223.
25. Yoo KH, et al. The impact of electronic medical records on timeliness of diagnosis of asthma. *J Asthma* 2007; 44(9): 753–758.
26. Ammenwerth E, et al. The effect of electronic prescribing on medication errors and adverse drug events: A systematic review. *J Am Med Inform Assoc* 2008; 15(5): 585–600.
27. Litwin AS. Not featherbedding but feathering the nest: Human resource management and investments in information technology. *Ind Relat*. forthcoming.
28. Ichniowski C, et al. What works at work: Overview and assessment. *Ind Relat* 1996; 35(3): 299–333.
29. Sittig DF, et al. Lessons from "Unexpected increased mortality after implementation of a commercially sold computerized physician order entry system". *Pediatrics* 2006; 118(2): 797–801.

30. Ash JS, et al. The extent and importance of unintended consequences related to computerized provider order entry. *J Am Med Inform Assoc* 2007; 14(4): 415–423.
31. Talmon J, et al. STARE-HI – statement on reporting of evaluation studies in health informatics. *Int J Med Inf* 2009; 78(1): 1–9.
32. Brynjolfsson E, Hitt LM, Yang S. Intangible assets: Computers and organizational capital. *Brookings Pap Eco Ac* 2002; 1: 137–181.
33. Poon EG, et al. Relationship between use of electronic health record features and health care quality: Results of a statewide survey. *Med Care* 2010; 48(3): 203–209.
34. Yu FB, et al. Full implementation of computerized physician order entry and medication-related quality outcomes: A study of 3364 hospitals. *Am J Med Qual* 2009; 24(4): 278–286.
35. McCullough JS, Casey M, Moscovice I, Prasad S. The effect of health information technology on quality in US hospitals. *Health Aff* 2010; 29(4): 647–654.
36. Devaraj S, Kohli R. Performance impacts of information technology: Is actual usage the missing link. *Manage Sci* 2003; 49(3): 273–289.
37. Keyhani S, et al. Electronic health record components and the quality of care. *Med Care* 2008; 46(12): 1267–1272.
38. Freeman RB. Longitudinal analyses of the effects of trade unions. *J Labor Econ* 1984; 2(1): 1–26.
39. Robinson JC, et al. Financial incentives, quality improvement programs, and the adoption of clinical information technology. *Med Care* 2009; 47(4): 411–417.
40. Hitt LM, Brynjolfsson E. Information technology and internal firm organization: An exploratory analysis. *J Manage Inform Syst* 1997; 14(2): 81–101.
41. Han YY, et al. Unexpected increased mortality after implementation of a commercially sold physician order entry system. *Pediatrics* 2005; 116(6): 1506–1512.
42. Eaton AE. The survival of employee participation programs in unionized settings. *Ind Labor Relat Rev* 1994; 47(3): 371–389.
43. Amarasingham R, et al. Clinical information technologies and inpatient outcomes: A multiple hospital study. *Arch Intern Med* 2009; 169(2): 108–114.
44. Skrondal A, Rabe-Hesketh S. Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models. Boca Raton: Chapman and Hall/CRC; 2004.
45. Huber GP, Power DJ. Retrospective reports of strategic-level managers: Guidelines for increasing their accuracy. *Strateg Manage J* 1985; 6(2): 171–180.
46. Gerhart B, Wright PM, MacMahan GC, Snell SA. Measurement error in research on human resources and firm performance: How much error is there and how does it influence effect size estimates? *Pers Psychol* 2000; 53(4): 803–834.
47. Kaushal R, et al. Electronic prescribing improves medication safety in community-based office practices. *J Gen Intern Med* 2010; 25(6): 530–536.
48. Buntin MB, Burke MF, Hoaglin MC, Blumenthal D. The benefits of health information technology: A review of the recent literature shows predominantly positive results. *Health Aff* 2011; 30(3): 464–471.