

---

## Research and Applications

# Does health informatics have a replication crisis?

Enrico Coiera,<sup>1</sup> Elske Ammenwerth,<sup>2</sup> Andrew Georgiou,<sup>1</sup> and Farah Magrabi<sup>1</sup>

<sup>1</sup>Australian Institute of Health Innovation, Macquarie University, NSW 2109, Australia and <sup>2</sup>University for Health Sciences, Medical Informatics and Technology, Austria

Corresponding Author: Enrico Coiera, Australian Institute of Health Innovation, Macquarie University, NSW 2109, Australia; enrico.coiera@mq.edu.au

Received 22 December 2017; Revised 25 February 2018; Editorial Decision 6 March 2018; Accepted 13 March 2018

### ABSTRACT

**Objective:** Many research fields, including psychology and basic medical sciences, struggle with poor reproducibility of reported studies. Biomedical and health informatics is unlikely to be immune to these challenges. This paper explores replication in informatics and the unique challenges the discipline faces.

**Methods:** Narrative review of recent literature on research replication challenges.

**Results:** While there is growing interest in re-analysis of existing data, experimental replication studies appear uncommon in informatics. Context effects are a particular challenge as they make ensuring replication fidelity difficult, and the same intervention will never quite reproduce the same result in different settings. Replication studies take many forms, trading-off testing validity of past findings against testing generalizability. Exact and partial replication designs emphasize testing validity while quasi and conceptual studies test generalizability of an underlying model or hypothesis with different methods or in a different setting.

**Conclusions:** The cost of poor replication is a weakening in the quality of published research and the evidence-based foundation of health informatics. The benefits of replication include increased rigor in research, and the development of evaluation methods that distinguish the impact of context and the nonreproducibility of research. Taking replication seriously is essential if biomedical and health informatics is to be an evidence-based discipline.

**Key words:** evaluation, replication, evidence-based informatics, research bias

---

### INTRODUCTION

The engine of science is designed according to some foundational principles, and foremost among these is the principle of falsifiability. We place greater faith in research that has faced repeated experimental challenges to prove it wrong.<sup>1</sup> Despite such foundations, it has been asserted that in the current research climate it is more likely for research claims to be false than true.<sup>2</sup>

The inability for researchers to reproduce many of the findings of past studies is causing particular concern in several disciplines, including psychology and the medical sciences. Indeed, one recent study suggests that “irreproducible preclinical research exceeds 50%” of all studies.<sup>3</sup> The poor reproducibility of experimental findings in any discipline can be due to problems with experimental design, statistical

errors, small sample sizes, outcome switching,<sup>4</sup> selective reporting of significant results (*p*-hacking),<sup>5</sup> failure to report negative results,<sup>6</sup> or journal publication biases which favor positive over negative results and “newsworthy” over confirmatory research.<sup>7,8</sup>

An antidote to these many and varied problems is to independently reproduce experimental results using a replication study. A replication study seeks to formally test whether an idea shown previously to be likely is in fact probable, or instead is the outcome of experimental or reporting problems. For studies in psychology, the “replication crisis” has led to a large international collaborative effort that attempted replications for 100 published experiments, and succeed in reproducing original results in only 40% of cases. Although somewhat controversial,<sup>9</sup> this massive project has led to the conclusion that a “large portion of replications produced weaker

evidence for the original findings despite using materials provided by the original authors.”<sup>10</sup>

The terms *replicate* and *reproduce* are used somewhat interchangeably in the literature, and will also be so used here. Reproduction is sometimes associated with the strict notion of exact copying of analyses on original data from a study to ensure that identical statistical results can be obtained in different hands,<sup>11</sup> (although some researchers have subsequently flipped the definitions of these 2 terms<sup>12</sup>).

Research in health and biomedical informatics is unlikely to be immune to these problems of research reproducibility. If anything, informatics is a discipline that emphasizes engineering, application, and “real world” testing over basic science. There is less of a tradition of controlled laboratory studies than in disciplines like psychology or the basic medical sciences, with a few exceptions such as usability or human–computer interaction studies. While randomized trials are sometimes undertaken in informatics, they are still not the norm,<sup>13</sup> and rarely do they seek to replicate earlier trials. Unlike disciplines such as pharmacology, informatics necessarily seeks often to embrace diversity in clinical settings and problems, rather than control for it.

We should thus anticipate that informatics must face its own challenge of research replication. Along the way we will need to not only tread the same path as other research disciplines, but also address our own unique set of challenges. As we do so, we will need to recognize that not all research can be replicated, and that not every replication is a wise use of scarce resources. Nor will all scientists accept that replication is a significant problem<sup>14</sup> nor that the benefits of replication outweigh the costs.<sup>15</sup>

The motivation for ensuring health informatics research is replicable is not just that we wish to see integrity in the research enterprise, and have confidence in the research evidence base. When informatics fails to properly evaluate new technologies, real world interventions or the hypotheses upon which they are based, the risks are at best that resources will be wasted in the futile deployment of technology, and that the quality, safety, and sustainability of the healthcare enterprise will not be enhanced. At worst, patients will be harmed.<sup>16</sup> A “replication crisis” in health informatics research would thus lead directly to an application crisis in its real-world implementation, and there is evidence enough that we do have problems in the application space.<sup>17</sup>

## CHALLENGES TO SCIENTIFIC REPLICATION IN HEALTH AND BIOMEDICAL INFORMATICS

When systematic reviews are undertaken in informatics, they often report that the studies included in the analysis exhibit wide heterogeneity in methods, settings, and in intervention, all making comparison between any two studies a challenge.<sup>18,19</sup> Direct experimental replication studies thus appear to be rare in health and biomedical informatics, and the causes for their relative absence are likely to be complex. Particular challenges that reduce the chance that a replication study will be conducted include difficulties in ensuring replication fidelity, the influence of study context, and the absence of a research culture that values replication.

### Replication Fidelity

When repeated studies are conducted in informatics, they often compare the same intervention but adapted to ensure it interoperates with the new local environment. This act of local adaption, however, means that we no longer are comparing similar interventions. This is even truer when the replication compares functionally similar

but independently developed software products. Small differences in design or engineering may introduce confounding factors that result in different outcomes. Similarly, variations in a bundle of different technology components that are implemented together, or variations in the strategy for their implementation including user training can all reduce replication fidelity. Even when observational studies use exactly the same methods but work off different population data sets, they can arrive at different conclusions.<sup>20</sup>

*Replication fidelity* is a measure of the similarity between the study methods and intervention used in a replication study and the original intervention. The less faithful a replication study is to the original study, the less we can rely on it to be a genuine test of the original study’s validity.

For example, the reporting of an apparent increased mortality associated with the introduction of computerized physician order entry (CPOE) was so controversial that it unusually triggered a number of follow on studies.<sup>21</sup> These new studies were attempted replications, in that they sought to retest the impact of CPOE on mortality, but the experiments took place at different institutions, with variations either in software, hardware, or implementation strategy.<sup>22–25</sup> Each “replication” was in effect a new experiment. Further, the new study protocols attempted to “fix” the perceived implementation problems of the original experiment through variations in implementation strategy. So the original study was never actually directly replicated.<sup>26</sup> The highest fidelity replication possible, repeating the original experiment at the same institution, on exactly the same system, and with the same implementation strategy, was not undertaken. So we are left still not knowing if the original result was valid.

Researchers in principle do have some flexibility in their choice of replication fidelity. Researchers are, however, limited by the resources it would take to achieve high fidelity, and by how well the original research is described. Even if the will and resources to undertake a replication study are present, if past research methods are poorly described, then accurate replication is not possible.

### The Cloud of Context

Much less under our experimental control is the *context* in which an experiment occurs. Introducing a high fidelity replication of an earlier informatics intervention in a different clinical setting, or the same setting at a later time, introduces a host of potentially confounding factors. These include variations in workflow, patient population and morbidity, resources, pre-existing infrastructure, and the education and experience of both clinical staff and patients. Such variations will also require an intervention to be customized to fit the new setting, and, thus, put a ceiling on the degree of experimental fidelity possible.

Indeed, the agreed wisdom in implementation science is that context effects in healthcare are so profound, that we should actually expect to see variations in outcome every time we repeat an intervention in a new setting.<sup>27</sup> In other words, this received wisdom suggests that by definition, differences in research outcome should be ascribed to changes in context, rather than a failure to replicate an earlier study. The logical conclusion of this argument is that replication is often simply not possible in informatics, that every experiment is unique and local, and that generalization from one experiment to another is limited at best, and not possible in many cases.

If that is genuinely true, then informatics has much more than a replicability crisis, but stands apart from most any other discipline that claims a scientific basis. We would be saying in effect that we are unable to separate statistically or methodologically aberrant results from robust and repeatable results. We would be arguing against the development of general principles and foundational theory that could

successfully and predictably guide the design and implementation of informatics interventions in the real world. Clearly this cannot be the case. Addressing the challenge of separating contextual variation from failure to replicate is, thus, a foundational challenge for the evaluation of health informatics interventions.

### Developing a Culture of Research Replication in Informatics

Along with the many technical challenges posed by replication, there are organizational and cultural challenges to recognizing and addressing the need for research reproducibility. The last decade has rightly seen a strong argument made for the critical importance of evaluation of information technology interventions.<sup>16</sup> Replication studies are an important pillar of research evaluation.

Informatics is not without positive examples of replication. The growing emphasis across all of the health sciences on registering study protocols, and depositing experimental data in a way that permits independent analysis will reduce barriers to validating study results. The creation of common data sets that allow multiple groups to independently analyze and share results<sup>28</sup> and of data mapping and analysis infrastructures such as Observational Health Data Sciences and Informatics<sup>29</sup> also help foster a culture of replication and validation among researchers. The different attempts to replicate the Han et al.<sup>21</sup> CPOE study demonstrate that experimental replication in different clinical contexts can play an important role in the science of health informatics, although such examples appear rarer.

There thus is a growing research culture around data sharing, but less so for experimental replication in different clinical contexts. Yet without such replication studies, there is no mechanism to test that a specific result in a specific location is robust enough to generalize to other settings. Without generalization, we find ourselves in a discipline that is unable to develop deeper theory that can guide our understanding and real world applications, and remains more focused instead on the local and the transient. Developing a culture that values replication should be seen as the essential next step in the evaluation journey.

### POSSIBLE APPROACHES TO THE DESIGN OF REPLICATION STUDIES IN HEALTH AND BIOMEDICAL INFORMATICS

Fortunately, informatics is not alone in having to deal with the challenge of replication fidelity and strong context effects. Biological disciplines such as ecology or evolutionary biology also have a long history of repeating studies using different species or ecosystems.<sup>15</sup> They too face similar challenges of generalization of lessons from individual studies. They also struggle with what it means to replicate research when high fidelity replication is not always possible.

The emerging thinking in the biological sciences is that we should take a portfolio approach to replication, using a variety of replication designs. The choice of replication design should be influenced by the research problem at hand, and the goal of the replication. In particular, there is a fundamental trade-off for researchers to make between testing the validity of a previous study and testing its generalizability. The more closely a study mirrors a previous study, the more it is a direct test of the validity of the original result. However, generalizability decreases as replication fidelity increases. Only when we test the same principle in different contexts or with different methods do we accumulate evidence of a principle's generalizability. This trade-off means that there is a spectrum of replica-

tion study designs one might choose from, depending on the replication goal. There appear to be at least 5 different forms of replication study<sup>15,30</sup> (Table 1):

- *Exact (or close) replication*: Such studies emphasize high fidelity replication of an earlier study, most possible in a controlled experimental laboratory setting. For example, a laboratory study of the usability of a specific CPOE system could be repeated in a different laboratory by different investigators using the same protocol and system. In real-world settings however, high fidelity replications will suffer from some variation in experiential context and so are more likely be “close” rather than exact replications. Indeed there is probably no such thing as a perfectly exact replication, as every replication study will differ from the original in some way.<sup>9</sup> The term “reproduction” is sometimes used to describe a strict copying of statistical analyses on original data, so that “executing the code on the data provided... produces results matching those that the authors claim”<sup>11</sup> and is the most narrow form of replication possible.
- *Partial replication*: Further along the spectrum from close replications, we allow the introduction of limited variation in research method, or the components of the intervention bundle. Such slight variations might be justified to allow replication in a different setting, or to fix obvious limitations with the earlier study, but should be small enough to allow the replication experiment to potentially reject the original finding. For example, a partial replication study could introduce the identical CPOE system used in an original study in a similar clinical environment, using the identical implementation strategy, and enrolling comparable groups of patients and clinicians.
- *Conceptual replication*: Sometimes we wish to employ entirely distinct tests of the same hypothesis, using very different study designs and even settings. The underlying hypothesis about a general underlying principle is the only thing shared in such a case. Similar results would provide evidence for the shared hypothesis. Differences in results might mean one of the studies was in error, or that both studies are correct in the context of their methods. For example, to test the hypothesis that all CPOE systems increase mortality rates, a conceptual study would trial a different version of a CPOE system to those used in other studies, and also vary implementation strategy, clinical setting, and research subjects.
- *Quasi replication (partial)*: When researchers wish to replicate a prior finding but also to extend it, we trade-off fidelity in the search for generalizable principles. Intentional variations in bundle or context<sup>31</sup> might seek to test the impact of the variation in comparison to the original results. For example, to test the impact of different implementation strategies on mortality rates after CPOE is introduced, a quasi replication study would take the same CPOE system that was used in an original study into a comparable setting, but use a different implementation strategy.
- *Quasi replication (conceptual)*: When a quasi-replication makes no effort to match earlier methods, the replication becomes more conceptual than direct. Such studies can help test the generality of prior results, but do not allow strong conclusions when results conflict. For example, inspired by evidence that CPOE use might be associated with mortality changes, researchers might next seek to test if this is a more widespread phenomenon, and examine mortality rates for different classes of information system, such as electronic health records.

Both validating studies (like exact and partial designs) and generalizing studies (such as quasi and conceptual) are needed in a

**Table 1.** Replication Studies Take Many Forms, Depending on the Fidelity of the Replication in Comparison to the Original, and the Hypothesis Being Tested, and Have Different Utility Depending on the Purpose of the Replication<sup>15</sup>

Replication study type	Example study	Utility of replication study design
Exact (or close) replication	A laboratory study of the usability of a specific CPOE system is repeated in a different laboratory using the exact same protocol and system	High fidelity replications test the validity of an earlier study
Partial replication	A clinical trial of a CPOE system is repeated using the same system in a similar clinical environment, using an identical implementation strategy, and enrolling comparable groups of patients and clinicians	Modest level fidelity replications test the validity of an earlier study when it is not possible to undertake high fidelity studies
Conceptual replication	Following a trial of a CPOE system in a clinical setting that shows mortality effects, the general hypothesis that all CPOE systems increase mortality rates is tested by using a different CPOE system, with a different implementation strategy, clinical setting and research subjects	Conceptual studies test the generalizability of past results, by sharing common hypotheses but using different clinical settings or methods
Quasi replication (partial)	To test the impact of implementation strategies on mortality rates after a particular CPOE is trialed, the same CPOE system is now tested in a comparable setting, but use a different implementation strategy	Quasi-replications seek to extend earlier experiments by including novel elements or hypotheses to build on the prior work, not just replicate it
Quasi replication (conceptual)	With evidence that CPOE use is associated with mortality changes, researchers test if this is generalizable to other system classes. They test the hypothesis that many clinical systems can affect mortality rates with an experiment using electronic health records and measuring mortality effects	The lowest fidelity form of replication, these studies help test the generality of prior results, but do not allow strong conclusions when their results conflict with earlier studies

discipline. Otherwise one could have a strong exact validation of a study, but fail to realize that the result does not replicate to other contexts because it is highly specific to the study protocol, or the study setting. Equally, triangulating a set of quasi or conceptual studies to identify some common generalization without first validating the primary studies may simply generate a false unification.

It is likely that what passes for replication in informatics is often some form of quasi-replication, where loosely similar interventions are trialed in loosely similar settings. Given the lack of formal thinking around replication study design, we may be assuming that these quasi replication studies are testing the validity of past studies (which they cannot) rather than testing (to a limited extent) their generalizability. Quasi replications cannot provide the same strength of evidence that exact replication does. When quasi-replicated studies differ in their results, we cannot separate a rejection of the original study results from appropriate variations in results due to a new study design or change in context.

Implementation science is slowly developing theoretical models of context that begin to allow modelling of the impact of context on outcome.<sup>27</sup> Unfortunately, these models for now remain qualitative, and we lack strong quantitative methods to appropriately adjust for known context effects. Nevertheless, it is standard practice in observational research to statistically adjust for the variations in populations that might lead to outcome variations. It should be no great stretch for quasi replication studies to identify important confounders associated with a change in context or intervention, and statistically model the effects of adjusting for them.

## A REPLICATION AGENDA FOR HEALTH AND BIOMEDICAL INFORMATICS

Many important consequences follow from this analysis. First and foremost, informatics researchers must recognize the primary importance of replication studies as a foundation for evidence-based

health informatics.<sup>16</sup> Replication needs to be done, done well, valued and published.

Secondly, unlike other major fields of research, we have yet to grapple with the implications that weak replication has for the integrity of our research knowledge base. We have yet to undertake any formal stock taking of the degree to which replication studies occur, and whether these are robust and meaningful. We have yet to consider the implications, as other disciplines have, of poor attention to replication. We might discover that many core research results in our discipline have not been properly tested, and among those are results that would fail the challenge of replication.

With the benefit of other disciplines already well on the road to reinstating replication as a core activity of the research community, it is possible to sketch out what the replication agenda looks like for informatics:

- *The role of replication:* It will be foundational to establish when replication studies are necessary and when further replications add little that is new to the health informatics evidence-base. Resources for research are always limited, and investigators will need to be able to identify which key study results require replication and which do not. Flags that trigger replication might include the theoretical or practical importance of a new result, the unexpectedness of a result in the context of what is already known, concerns about the conduct of a research study, or study weaknesses such as small effect and sample sizes. The balance between conducting replications that validate a result, and studies that generalize a result by testing its underlying hypothesis will also depend on the feasibility of validation and the existence of other validation studies. Primary studies that have already been independently validated, or that provide small increments to an existing knowledge base are likely to have a lower priority.
- *The role of journals:* It will be important to publish important replication studies, but equally to avoid flooding the literature with low value studies that tell us nothing new. Journals typically



are biased to rewarding novelty, and replications that do not upturn an original result but instead confirm it are more likely to be seen as not novel. Yet in scientific terms there is no basis for such a publication bias, at least for the first replication of an important result. It is already common for journals to create companion publications devoted to research protocols, which are seen as scientifically important but not necessarily “newsworthy.” A similar case might be made for creating companion journals for scientifically important, but less novel replications.

- *Revised standards for primary studies to enhance replication:* Replication is only possible when there is a clear template to follow from the study that is being replicated. If the informatics community is to take replication seriously, then we will need to examine how well informatics interventions and their implementation strategy are described in study protocols, and devote effort to agreeing what are the essential elements in a protocol from a reproducibility perspective. The Statement on the Reporting of Evaluation Studies in Health Informatics (STARE-HI) guidelines are an important step in this direction, as they define a minimum content set for health IT evaluation publications.<sup>32</sup> Current initiatives that encourage or mandate the publication of research protocols ahead of reporting results should provide greater clarity about the methods used for a study.<sup>33</sup> Encouraging the release of study data allows other scientists to scrutinize a study, checking for integrity of the primary data, and carry out essential replication of statistical analyses.<sup>34</sup> Conflict of interest registries also allow scrutiny of investigators and their motivations for conducting a study.<sup>35</sup>
- *Standards and tools for replication studies:* Best practice standards for replication studies are needed, including guidance on which form of replication study is most appropriate for a particular research question, on mechanisms for maximizing interventional fidelity, and for drawing appropriate conclusions when making comparisons between studies. Creation of the infrastructure for data commons, and for sharing of analytic tools are in progress,<sup>29</sup> and their maturation will make it much easier to undertake direct analytic validations of results from primary research data, as well as support the re-use of common methods across different experiments.<sup>36</sup>
- *Contextual reporting and analysis:* Coming to terms with context is a foundational challenge for health informatics. Being able to describe context explicitly will assist both in the task of replication, by quantifying study comparability, as well as in the translation of research evidence into real world application. The long-term goal must be to develop methods that allow determination of the likelihood that study results in one setting can transfer to a different clinical setting, population or organization. The best way to describe the context of a study to allow generalization to other settings remains an open question in implementation science. It should be possible for a consensus set of context variables and associated scores to be agreed upon when reporting studies. For example, it should be possible to describe the local context variables most likely to influence the adoption of an electronic record or decision support system in a hospital setting. Consensus processes and automated methods can assist in defining such important indicator variables.<sup>37</sup> Health services researchers are slowly developing standardized templates to encourage more uniform reporting of context.<sup>38</sup> The informatics community thus needs to engage with such implementation science and bring theoretically sound models of context influence into study design, conduct, and analysis.
- *Peer review:* There currently are no formal standards for peer review or for formal quality assessment of peer reviews in informat-

ics. Yet peer reviews must themselves be replicable, consistent, and identify issues, not just with statistical methods, but issues that would make replication of a study difficult. Studies with low replicability because of the conduct or reporting of the study need to be identified, and these weaknesses remedied prior to publication. Equally, reviewers will need to know how to assess a replication study. Significant effort has been devoted to the issue of improving the quality of peer review, and among the tools suggested are reviewer guidelines, structured review templates, and formal training before researchers are qualified to review.<sup>39,40</sup>

- *Replicating major results:* The consequence of ignoring replication is that there is doubt about the integrity of the research base for a discipline. To estimate the replicability of their research base, psychologists replicated 100 studies drawn from the published research in a single year.<sup>10</sup> Informatics researchers might be wise to do something similar. We could also randomly select studies from a defined period, or instead seek community input on key informatics results that require replication. That exercise would teach us much about the state of replication research in informatics, as well as provide an estimate of the quality of the informatics research evidence by identifying past results that require reconsideration or discarding.
- *Addressing cultural barriers:* Education in the evaluation of health informatics interventions remains a priority for the discipline.<sup>27</sup> There will always be a tension between the need for scientific rigor, and the often urgent needs for local systems to be implemented in a nonevidence based way, for example, because of local pressures, priorities, and context differences. It is easy to say that “we do things differently here,” and it is much harder to slow a project down and add cost to ensure both that past lessons are heeded and that others may learn from the exercise in the future. Addressing the scientific risks that come with a failure to replicate a primary evaluation thus is a significant cultural challenge. Embedding replication thinking into evaluation training is a good first step, but will not make much headway without leadership and support from champions for change within the discipline, and from those who would benefit from the contributions that informatics can make to healthcare delivery.

## CONCLUSION

Given the significant impact that informatics interventions can have on clinical processes and patient outcomes, it is surprising that replication studies are not a standard in informatics research. We live in an age where there is intense competition among researchers to publish in high-ranking journals, and the result is a pressure to publish quickly, and to emphasize novelty. The rock stars of research are the ones who publish new, potentially discipline challenging studies, less so those who come in afterwards to test the validity of the initial reports. Incentives are biased against replication. The cost, as many disciplines have now discovered, is a significant weakening in the quality of published research, and doubts about the validity of their evidence base.

Informatics research is highly unlikely to be different. We additionally face challenges in the design and execution of replication studies. We have taken as a mantra that different outcomes between similar studies are the consequence of context and implementation changes. We much less often consider the obvious alternative, that failure to replicate may mean the original study was flawed. Learning to separate these effects will in of itself be a new research challenge, and may lead us to a deeper, richer, more theoretically robust, understanding of informatics and the nature of digital interventions in a complex socio-

technical universe. It will also mean that real world applications of health informatics research, targeted at improving the quality, safety, and efficiency of healthcare, have a much greater chance of success.

## FUNDING

This research was supported by the National Health and Medical Research Council (NHMRC) grant APP1134919 (Centre for Research Excellence in Digital Health).

## COMPETING INTERESTS

None.

## CONTRIBUTORS

EC conceived the study and wrote the initial draft. EA, FM, and AG all participated in the design of the analysis, critical review, and writing of the final text. All authors approved the final draft.

## REFERENCES

- Popper K. *Conjectures and Refutations*. London: Routledge and Kegan Paul; 1963.
- Ioannidis JP. Why most published research findings are false. *PLoS Med* 2005;2(8):e124.
- Freedman LP, Cockburn IM, Simcoe TS. The economics of reproducibility in preclinical research. *PLoS Biol* 2015;13(6):e1002165.
- Mathieu S, Boutron I, Moher D, Altman DG, Ravaud P. Comparison of registered and published primary outcomes in randomized controlled trials. *JAMA* 2009;302(9):977–984.
- Simonsohn U, Simmons JP, Nelson LD. Better P-curves: making P-curve analysis more robust to errors, fraud, and ambitious P-hacking, a Reply to Ulrich and Miller. *J Exp Psychol Gen* 2015;144(6):1146–1152.
- Chalmers I. Underreporting research is scientific misconduct. *JAMA* 1990;263(10):1405–1408.
- Macleod MR, McLean AL, Kyriakopoulou A, et al. Risk of bias in reports of in vivo research: a focus for improvement. *PLoS Biol* 2015;13(10):e1002273.
- Curtis MJ, Abernethy DR. Replication – why we need to publish our findings. *Pharmacol Res Perspect* 2015;3(4):e00164.
- Anderson CJ, Bahník Š, Barnett-Cowan M, et al. Response to Comment on “Estimating the reproducibility of psychological science”. *Science* 2016;351:1037–1039.
- Open Science Collaboration. Estimating the reproducibility of psychological science. *Science* 2015;349(6251):943.
- Peng RD. Reproducible research and Biostatistics. *Biostatistics* 2009;10(3):405–408.
- Drummond C. Replicability is not reproducibility: nor is it good science. In: Proceedings of the Evaluation Methods for Machine Learning Workshop 26th International Conference for Machine Learning; 2009:4; Montreal.
- Liu JLY, Wyatt JC. The case for randomized controlled trials to assess the impact of clinical information systems. *J Am Med Inform Assoc* 2011;18(2):173–180.
- Pashler H, Harris CR. Is the replicability crisis overblown? Three arguments examined. *Perspect Psychol Sci* 2012;7(6):531–536.
- Nakagawa S, Parker TH. Replicating research in ecology and evolution: feasibility, incentives, and the cost-benefit conundrum. *BMC Biol* 2015;13(1):88.
- Ammenwerth E, Rigby M. *Evidence-Based Health Informatics*. Amsterdam: IOS Press; 2016.
- Kim MO, Coiera E, Magrabi F. Problems with health information technology and their effects on care delivery and patient outcomes: a systematic review. *J Am Med Inform Assoc* 2017;24(2):246–250.
- Black AD, Car J, Pagliari C, et al. The impact of eHealth on the quality and safety of health care: a systematic overview. *PLoS Med* 2011;8(1):e1000387.
- Shekelle P, Morton SC, Keeler EB. Costs and Benefits of Health Information Technology. *Evidence Report/Technology Assessment*. Rockville, MD: AHRQ; 2006.
- Madigan D, Ryan PB, Schuemie M, et al. Evaluating the impact of database heterogeneity on observational study results. *Am J Epidemiol* 2013;178(4):645–651.
- Han YY, Carcillo JA, Venkataraman ST, et al. Unexpected increased mortality after implementation of a commercially sold computerized physician order entry system. *Pediatrics* 2005;116(6):1506–1512.
- Brunette DD, Tersteeg J, Brown N, et al. Implementation of computerized physician order entry for critical patients in an Academic Emergency Department is not associated with a change in mortality rate. *Western J Emerg Med* 2013;14(2):114–120.
- Al-Dorzi HM, Tamim HM, Cherfan A, Hassan MA, Taher S, Arabi YM. Impact of computerized physician order entry (CPOE) system on the outcome of critically ill adult patients: a before-after study. *BMC Med Inform Decis Making* 2011;11:71.
- Del Beccaro MA, Jeffries HE, Eisenberg MA, Harry ED. Computerized provider order entry implementation: no association with increased mortality rates in an intensive care unit. *Pediatrics* 2006;118(1):290–295.
- Longhurst CA, Parast L, Sandborg CI, et al. Decrease in hospital-wide mortality rate after implementation of a commercially sold computerized physician order entry system. *Pediatrics* 2010;126(1):14–21.
- Ammenwerth E, Talmon J, Ash JS, et al. Impact of CPOE on mortality rates – contradictory findings, important messages. *Methods Archive* 2006;45(6):586–593.
- Helfrich CD, Damschroder LJ, Hagedorn HJ, et al. A critical synthesis of literature on the promoting action on research implementation in health services (PARIHS) framework. *Implement Sci* 2010;5:82.
- Johnson AEW, Pollard TJ, Shen L, et al. MIMIC-III, a freely accessible critical care database. *Scientific Data* 2016;3:160035.
- Hripsak G, Duke JD, Shah NH, et al. Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers. *Stud Health Technol Inform* 2015;216:574–578.
- Kelly CD. Replicating empirical research in behavioral ecology: how and why it should be done but rarely ever is. *Quart Rev Biol* 2006;81(3):221–236.
- Palmer AR. Quasi-replication and the contract of error: lessons from sex ratios, heritabilities and fluctuating asymmetry. *Ann Rev Ecol Syst* 2000;31(1):441–480.
- Brender J, Talmon J, de Keizer N, Nykänen P, Rigby M, Ammenwerth E. STARE-HI – Statement on Reporting of Evaluation Studies in Health Informatics: Explanation and Elaboration. *Appl Clin Inform* 2013;4(3):331–358.
- De Angelis C, Drazen JM, Frizelle FA, et al. Clinical trial registration: a statement from the International Committee of Medical Journal Editors. *New Engl J Med* 2004;351:1250–1251.
- Taichman DB, Sahni P, Pinborg A, et al. Data Sharing Statements for Clinical Trials—A Requirement of the International Committee of Medical Journal Editors. *N Engl J Med* 2017;376:2277–2279.
- Dunn AG, Coiera E, Bourgeois FT, Mandl KD. Conflict of interest disclosure in biomedical research: a review of current practices, biases, and the role of public registries in improving transparency. *Res Integrity Peer Rev* 2016;1(1):1.
- Almugbel R, Hung L-H, Hu J, et al. Reproducible Bioconductor workflows using browser-based interactive notebooks and containers. *J Am Med Inform Assoc* 2018;25(1):4–12.
- Coiera E, Choong MK, Tsafnat G, Hibbert P, Runciman WB. Linking quality indicators to clinical trials: an automated approach. *Int J Qual Healthcare* 2017;29(4):571–578.
- Cotterill S, Knowles S, Martindale A-M, et al. Getting messier with TIDieR: embracing context and complexity in intervention reporting. *BMC Med Res Methodol* 2018;18(1):12.
- Walker R, Rocha da Silva P. Emerging trends in peer review—a survey. *Front Neurosci* 2015;9(169). doi: 10.3389/fnins.2015.00169.
- Moher D, Altman DG. Four proposals to help improve the medical research literature. *PLoS Med* 2015;12(9):e1001864.