

Is there a 'best measure' of patient safety?

Ann M Borzecki,^{1,2,3} Amy K Rosen^{1,4,5}

¹Center for Healthcare Organization and Implementation Research, Bedford VAMC Campus, Bedford, Massachusetts, USA

²Department of Health Law, Policy and Management, Boston University School of Public Health, Boston, Massachusetts, USA

³Department of Internal Medicine, Boston University School of Medicine, Boston, Massachusetts, USA

⁴Department of Surgery, Boston University School of Medicine, Boston, Massachusetts, USA

⁵Center for Healthcare Organization and Implementation Research, VA Boston Healthcare System Campus, Boston, Massachusetts, USA

Correspondence to

Dr Ann M Borzecki, Center for Healthcare Organization and Implementation Research, Bedford MA 01730, USA; amb@bu.edu

Accepted 25 September 2019



► <http://dx.doi.org/10.1136/bmjqs-2018-008852>



© Author(s) (or their employer(s)) 2019. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: Borzecki AM, Rosen AK. *BMJ Qual Saf* Epub ahead of print: [please include Day Month Year]. doi:10.1136/bmjqs-2019-009730

Despite consensus that preventing patient safety events is important, measurement of safety events remains challenging. This is, in part, because they occur relatively infrequently and are not always preventable. There is also no consensus on the 'best way' or the 'best measure' of patient safety. The purpose of all safety measures is to improve care and prevent safety events; this can be achieved by different means. If the overall goal of measuring patient safety is to capture the universe of safety events that occur, then broader measures encompassing large populations, such as those based on administrative data, may be preferable. Acknowledging the trade-off between comprehensiveness and accuracy, such measures may be better suited for surveillance and quality improvement (QI), rather than public reporting/reimbursement. Conversely, using measures for public reporting and pay-for-performance requires more narrowly focused measures that favour accuracy over comprehensiveness, such as those with restricted denominators or those based on medical record review.

There are at least two well-established patient safety measurement systems available for use in the inpatient setting, namely the administrative data-based Agency for Healthcare Research and Quality (AHRQ) Patient Safety Indicators (PSIs) and the medical record-based National Surgical Quality Improvement Programme (NSQIP) measures.¹⁻³ The AHRQ PSIs, publicly released in 2003, are evidence-based measures designed to screen for potentially preventable medical and surgical complications that occur in the acute care setting. Since they use administrative data, they were originally designed as tools for use in case finding for local QI efforts and surveillance, as well as for internal hospital comparisons. They were developed using a rigorous process beginning with a thorough review of the

literature for existing administrative data-based indicators, review by clinical expert panels, consultation with coding experts and empirical analyses to assess the statistical properties of the measures, such as reliability and predictive and construct validity. They were intentionally developed to favour specificity over sensitivity. As such, the indicators have various exclusion criteria designed to decrease the likelihood of including patients in whom a complication is very unlikely to be preventable, as well as having a well-specified denominator. (Prior to the inclusion of Present on Admission (POA) coding in US claims datasets in late 2007, patients with various secondary diagnoses were also excluded.)⁴ Additionally, to better compare hospitals and reflect the fact that some patients are at higher risk for a complication than others, risk adjustment is used to compare indicator rates.

The 'flagship' medical record or chart-based system, NSQIP, was established by the US Veterans Health Administration (VA) in 1994 over concerns of higher mortality rates and substandard surgical care in the VA. It was designed to promote QI of VA surgical care by providing reliable, valid and comparative information regarding 30-day surgical outcomes, such as morbidity and mortality, across all facilities performing major non-cardiac surgery.¹ Trained nurse reviewers prospectively gather medical record information from a select sample of all eligible operations.¹ The programme's success led to establishment and launch of a similar programme in the non-VA setting by the American College of Surgery (ACS) in 2004, known as ACS-NSQIP.⁵ Use of this programme has been expanding; it is currently being implemented in at least nine countries, including Canada, for benchmarking and QI purposes.⁵

In the current issue, McIsaac *et al* examined the accuracy of a new set of

administrative data-based PSIs developed using Canadian International Classification of Diseases (ICD)-10 coded data.^{6,7} The 'new' PSIs identify complications of care that arise after admission using a diagnosis timing variable ('diagnosis type') present in the administrative data which indicates whether the diagnosis is pre-existing or occurred after admission (potentially representing a complication of care). They were designed to improve on some of the recognised limitations of the AHRQ PSIs. Namely, they were designed to be more comprehensive (covering more complications) and applicable to a larger population of patients (they do not exclude populations at higher risk of a complication) than the AHRQ PSIs.⁶ Unlike the AHRQ PSIs, they do not specifically try to account for the potential preventability of an event; instead, they look for administrative data-based codes that may represent suboptimal quality or unsafe care. They also use a global denominator but they can be applied to a particular population of interest. Additionally, while the AHRQ PSIs are risk-adjusted, the new PSIs are not; only observed rates are calculated. Finally, the new PSIs were created through a slightly different process than the AHRQ PSIs.⁶ Rather than starting with the literature in the area, the developers identified all administrative codes representing conditions arising after admission.⁶ These individual codes were then rated by patient safety experts with respect to their likelihood of being related to a patient safety event and then grouped into categories (eg, hospital-acquired infections). These categories were not mutually exclusive such that codes could be assigned to more than one category.

McIsaac *et al* compared the accuracy of events identified by the new PSIs to those identified by NSQIP (considered the 'gold standard').⁷ Only 7 of 17 PSI categories mapped to a specific NSQIP complication and in some cases, a given NSQIP complication was mapped to more than one PSI. Of the overlapping complication categories, overall, they found low to moderate positive predictive values (PPVs) and sensitivities, and high specificities and negative predictive values of the new PSIs compared with NSQIP. Complications with the worst agreement included those related to fluid management and respiratory issues. With respect to the specific comparisons, among the overlapping categories, they did not report which specific events the PSIs identified that NSQIP did not. However, of the non-overlapping categories, the PSIs picked up several gastrointestinal complications while NSQIP identified events such as wound disruption which the new PSIs did not.

As the authors note, prior investigators have compared the AHRQ PSIs to NSQIP data and similarly found generally low sensitivities and low to moderate PPVs.⁸⁻¹⁰ These two previous studies used the AHRQ PSI measures based on ICD-9 codes; Romano *et al* lacked POA data while Cima *et al* incorporated POA

data into their study.^{8,9} Although we are not aware of any similar validation studies using the ICD-10 version of the AHRQ PSIs (ICD-10 is considered to be a more specific diagnosis-based coding system than ICD-9), we know from this prior work that an important reason for the seemingly low accuracy of the PSIs has to do with fundamental differences in definitions of medical record-based versus administrative-based safety events.⁹⁻¹¹ Despite overlap conceptually, PSIs (both the new and old) and NSQIP measures have different definitions that reflect the method of development as well as the data sources used. For example, the AHRQ PSI perioperative haemorrhage or haematoma (previously known as 'postoperative') gets mapped to the fairly specific NSQIP complication of postoperative bleeding requiring transfusion of greater than or equal to 4 units of blood. In the new PSI set, 'haemorrhage' gets mapped to this complication along with postoperative stroke which can be haemorrhagic or ischaemic.¹² Despite the difficulty in mapping NSQIP complications to PSI events, even in the validation work done by both AHRQ and our VA group, in which we used medical records as the 'gold standard' to validate diagnosis codes, coding system limitations (especially lack of POA codes and coding specificity, eg, many codes associated with complications were not specific with respect to timing and could be used for preoperative or postoperative events) as well as coding errors, accounted for many of the false positives; similar problems contributed to a large percentage of false negatives.¹³⁻¹⁵ Although our groups considered relatively few cases to represent documentation errors, other researchers have found that documentation quality affects coding accuracy.¹⁶

Notably, these prior PSI validation studies led to modifications of the ICD-9 coding system to improve the specificity of codes relevant to several of the AHRQ PSIs. A more recent study of the AHRQ PSI Postoperative Deep Vein Thrombosis and Pulmonary Embolism found that inclusion of POA data and the presence of more specific ICD-9 codes resulted in improved PPVs, from 43%–48% to 81%–99%.¹⁷ Although no US studies have examined the validity of the ICD-10 version of the AHRQ PSIs, one study using an ICD-10 based international version of the AHRQ PSIs with POA coding examined five PSIs and found fairly high PPVs for four of them (62.5%–86%).¹⁸

How does the study by McIsaac *et al* contribute to the existing literature?⁷ We think an important contribution of this study is that it demonstrates that even with both the more specific ICD-10-based measurement system and the equivalent of POA coding (diagnosis timing), the new PSI measures still suffer some of the same limitations as the AHRQ PSIs that originally used an ICD-9-based system without POA (which was added later to improve criterion validity). While this new PSI system also has the potential advantage of enhancing our ability to measure the universe of safety

events (ie, the new PSIs are broader with respect to the numerator and denominator compared with AHRQs, thereby identifying more events), without knowing more about the specifics of the events, it is hard to determine whether they are true positives and/or if they represent potentially preventable events. Use of a global denominator may capture many events that are not preventable due to patient-related or procedure-related factors. Furthermore, only three countries currently have datasets with timing diagnoses, so international use of such a system is relatively limited.

We were surprised at the high rate of complications identified by both the new PSIs and NSQIP (18.7% and 22%, respectively). Although this does not negate the authors' findings, these rates are higher than those of other studies. For example, Cima *et al* reported a complication rate of 7.4% using NSQIP data at one US hospital; Mull *et al* found a rate of 6% in a national VA sample.^{8 10} The current study's findings are based on one hospital network comprising two hospitals. Presumably, the accuracy of administrative data may vary by institution (as well as condition). The authors only examined surgical patients; would the new PSIs perform better or worse in medical patients?

So is there a best measure of patient safety? All measures have their strengths and weaknesses. Although the AHRQ PSIs were originally designed for QI and surveillance, in the USA, they have been increasingly used for federal and state public reporting and pay-for-performance despite concerns about coding accuracy.^{19 20} Differences in complications rates across sites could therefore reflect coding and documentation differences between facilities, rather than true differences in complication rates. The new PSIs may be broader in scope but appear to have similar limitations to the AHRQ ones when it comes to accuracy. The NSQIP-based system, on the other hand, has the advantage of high accuracy, as its measures are based on detailed clinical information; however, such measures are resource-intensive to collect and, due to sampling issues, only capture a small subset of post-surgical events.²¹ However, it is not necessary to choose one measure or set of measures over another. Having different measures of patient safety, including those using administrative data, contributes to the goal of comprehensive measurement. We think patient safety improvement may be best served by considering the different measurement systems as complementary, which will improve our ability to capture as many safety events as possible. Such measurement systems, even when used together, are not 'perfect', and thus would be most useful if they had a primary focus on QI use, rather than on public reporting or financial reimbursements, potentially punitive actions that may not be equitable across hospitals. Furthermore, since they are focused only on *inpatient* safety events, neither the new PSIs, AHRQ PSIs or NSQIP allow us to truly capture the universe of safety events. Given that most

care is now delivered in the outpatient setting, the field of patient safety measurement needs to expand to capture this setting as well. Continuing to 'reinvent' the wheel with development of new inpatient PSIs—that in the end have some of the same limitations as the older PSI measures—is a commendable journey, but one not likely to significantly advance the patient safety field. It is time to take the road less travelled.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Commissioned; internally peer reviewed.

REFERENCES

- 1 Khuri SF, Daley J, Henderson W, *et al*. The Department of Veterans Affairs' NSQIP: the first national, validated, outcome-based, risk-adjusted, and peer-controlled program for the measurement and enhancement of the quality of surgical care. National Va surgical quality improvement program. *Ann Surg* 1998;228:491–507.
- 2 McDonald KM, Romano PS, Geppert J. Technical Review 5 (Prepared by the UCSF-Stanford Evidence-Based Practice Center under Contract # 290-97-0013). AHRQ Pub. No. 02-0038. In: *Agency for health care research and quality*. Rockville, MD: Stanford, 2002.
- 3 Agency for Healthcare Research and Quality. Patient safety indicators overview. Available: https://www.qualityindicators.ahrq.gov/Modules/psi_resources.aspx [Accessed 10 Aug 2019].
- 4 CMS.gov. hospital-acquired conditions (present on admission indicator). Available: <https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/HospitalAcqCond/index.html?redirect=/hospitalacqcond> [Accessed 6 Aug 2019].
- 5 American College of Surgeons. ACS NSQIP - History. Available: <https://www.facs.org/quality-programs/acs-nsqip/about/history> [Accessed 10 Aug 2019].
- 6 Southern DA, Burnand B, Droesler SE, *et al*. Deriving ICD-10 codes for patient safety indicators for large-scale surveillance using administrative hospital data. *Med Care* 2017;55:252–60.
- 7 McIsaac DI, Hamilton GM, Abdulla K, *et al*. Validation of new ICD-10-based patient safety indicators for identification of in-hospital complications in surgical patients: a study of diagnostic accuracy. *BMJ Qual Saf* 2019;pii: bmjqs-2018-008852.
- 8 Cima RR, Lackore KA, Nehring SA, *et al*. How best to measure surgical quality? comparison of the agency for healthcare research and quality patient safety indicators (AHRQ-PSI) and the American College of surgeons national surgical quality improvement program (ACS-NSQIP) postoperative adverse events at a single institution. *Surgery* 2011;150:943–9.
- 9 Romano PS, Mull HJ, Rivard PE, *et al*. Validity of selected AHRQ patient safety indicators based on Va national surgical quality improvement program data. *Health Serv Res* 2009;44:182–204.
- 10 Mull HJ, Borzecki AM, Loveland S, *et al*. Detecting adverse events in surgery: comparing events detected by the Veterans health administration surgical quality improvement program and the patient safety indicators. *Am J Surg* 2014;207:584–95.

- 11 Koch CG, Li L, Hixson E, *et al.* What are the real rates of postoperative complications: elucidating inconsistencies between administrative and clinical data sources. *J Am Coll Surg* 2012;214:798–805.
- 12 Agency for Healthcare Research and Quality. Patient Safety Indicator Resources - Technical Specifications. Available: https://www.qualityindicators.ahrq.gov/Modules/PSI_TechSpec_ICD10_v2019.aspx [Accessed 11 Aug 2019].
- 13 Rosen AK, Itani KMF, Cevasco M, *et al.* Validating the patient safety indicators in the Veterans health administration: do they accurately identify true safety events? *Med Care* 2012;50:74–85.
- 14 Borzecki AM, Cevasco M, Chen Q, *et al.* Improving identification of postoperative respiratory failure missed by the patient safety indicator algorithm. *Am J Med Qual* 2013;28:315–23.
- 15 Borzecki AM, Cevasco M, Mull H, *et al.* Improving the identification of postoperative wound dehiscence missed by the patient safety indicator algorithm. *Am J Surg* 2013;205:674–80.
- 16 So L, Beck CA, Brien S, *et al.* Chart documentation quality and its relationship to the validity of administrative data discharge records. *Health Informatics J* 2010;16:101–13.
- 17 Sadeghi B, White RH, Maynard G, *et al.* Improved coding of postoperative deep vein thrombosis and pulmonary embolism in administrative data (AHRQ patient safety indicator 12) after introduction of new ICD-9-CM diagnosis codes. *Med Care* 2015;53:e37–40.
- 18 Quan H, Eastwood C, Cunningham CT, *et al.* Validity of AHRQ patient safety indicators derived from ICD-10 hospital discharge Abstract data (chart review study). *BMJ Open* 2013;3:e003716.
- 19 U.S. Centers for Medicare & Medicaid Services. Hospital compare. Available: <http://www.medicare.gov/hospitalcompare/?AspxAutoDetectCookieSupport=1> [Accessed 15 Aug 2019].
- 20 Hospital Inpatient Quality Reporting (IQR). Program overview: resources for fiscal year 2012 payment determination. program changes. Available: <http://www.qualitynet.org/dcs/ContentServer?cid=1138115987129&pagename=QnetPublic%2FPage%2FQnetTier2&c=Page> [Accessed 10 Aug 2019].
- 21 Henderson WG, Daley J. Design and statistical methodology of the National surgical quality improvement program: why is it what it is? *Am J Surg* 2009;198(5 Suppl):S19–S27.