International Journal for Quality in Health Care, 2021, 30(1), 1–6
doi: 10.1093/intqhc/mzab025
Advance Access Publication Date: 5 February 2021
Original Research Article

Original Research Article

Mitigating imperfect data validity in administrative data PSIs: a method for estimating true adverse event rates

BASTIEN BOUSSAT[1,2,3], HUDE QUAN[1], JOSE LABARERE[2,3], DANIELLE SOUTHERN[1], CHANTAL M. COURIS[4], and WILLIAM A. GHALI[1]

[1]Department of Community Health Sciences, Cumming School of Medicine, O'Brien Institute for Public Health, University of Calgary, TRW Building, 3280 Hospital Drive NW, Calgary, AB T2N 1N4, Canada, [2]Quality of Care Unit, Grenoble University Hospital, Boulevard de la Chantourne, 38043 cedex 09, Grenoble, France, [3]TIMC UMR 5525 CNRS, Computational and Mathematical Biology Team, Grenoble Alpes University, Boulevard de la Chantourne, Pavillon Taillefer, 38043 cedex 09, Grenoble, France, and [4]Canadian Institute for Health Information, Indicator Research and Development Team, Research and Analysis Division, 4110 Yonge Street, Suite 300, Toronto, ON M2P 2B7, Canada

Address reprint requests to: Bastien Boussat, Department of Community Health Sciences, Center for Health Informatics, Cumming School of Medicine, University of Calgary, TRW Building, 3280 Hospital Drive NW, Calgary, AB T2N 4Z6, Canada. Tel: +1 (403) 210-8996; Fax: +1 (403) 270-7307; E-mail: bastien.boussat@ucalgary.ca

Received 17 June 2020; Editorial Decision 24 January 2021; Revised 22 November 2020; Accepted 4 February 2021

Abstract

Question: Are there ways to mitigate the challenges associated with imperfect data validity in Patient Safety Indicator (PSI) report cards?
Findings: Applying a methodological framework on simulated PSI report card data, we compare the adjusted PSI rates of three hospitals with variable quality of data and coding. This framework combines (i) a measure of PSI rates using existing algorithms; (ii) a medical record review on a small random sample of charts to produce a measure of hospital-specific data validity and (iii) a simple Bayesian calculation to derive estimated true PSI rates. For example, the estimated true PSI rate, for a theoretical hospital with a moderately good quality of coding, could be three times as high as the measured rate (for example, 1.4% rather than 0.5%). For a theoretical hospital with relatively poor quality of coding, the difference could be 50-fold (for example, 5.0% rather than 0.1%).
Meaning: Combining a medical chart review on a limited number of medical charts at the hospital level creates an approach to producing health system report cards with estimates of true hospital-level adverse event rates.

Key words: Patient safety; Adverse event; Patient safety indicators; Administrative data; Bayesian adjustment

Introduction

In the USA, the Centers for Medicare & Medicaid Service (CMS) began using selected Agency for Healthcare Research and Quality (AHRQ) Patient Safety Indicators (PSI-90) as a core metric in pay-for-performance programs on 1 October 2014 [1]. In the Hospital-Acquired Condition Reduction Program, PSI-90 accounts for 15% of the overall score, and the poorest-performing hospitals with Total HAC Scores greater than the 75th percentile have their CSM payments reduced by 1%. However, numerous flaws have been reported by the National Quality Forum, including an inaccuracy of adverse events identified. In Europe, most pay-for-performance programs also include quality and safety indicators as a core metric and are therefore exposed to similar flaws [2–5]. As reported by Rajaram *et al.*, there is concern that flaws relating to PSIs may incorrectly flag

problem areas, unfairly penalize hospitals financially and adversely influence clinician engagement in quality improvement [6]. Adverse event data are not perfect, and this can produce misleading PSI rates.

Measuring and reporting on quality in health care

Measuring and reporting on the quality of health care has become fundamental in recent years. Indeed, quality inexorably supplants volume as the main performance index in the health sector. The public now expects real transparency, and health organizations are responding by publicly reporting the performance of hospitals and physicians [2, 3, 7–9]. While real and laudable efforts are being made to provide this information, methods measuring the quality of health care remain imperfect and do not allow the public to determine whether the measurements and data are accurate and transparent.

No perfect method

No perfect method exists to achieve this goal. Self-reporting of adverse events is well promoted and implemented at the hospital level but underestimates adverse events [10]. Furthermore, self-reporting is time consuming for busy frontline staff and is impaired by complex sociological factors such as fear of punishment, shame and lack of education. Voluntary reporting has undeniable benefits, as it involves health-care providers in an active process of self-reflection in the management of patient safety, but is not a gold standard for measuring and reporting on rates of patient harm.

The manual review of medical records is the most commonly used method for epidemiologic purposes [11]. This method is more accurate, and it permits attempts at distinguishing preventable and unpreventable adverse events. This approach also has clear weaknesses; reviewing medical charts is laborious, requires some level of medical knowledge and has imperfect inter-rater reliability, raising questions about feasibility, sustainability and opportunity costs.

Other methods using automated data abstraction from administrative data are now widely used [12]. Algorithms derived from the International Classification of Disease (ICD) codes allow reporting on indicators for patient safety. However, administrative data are produced manually to document the patients' characteristics and care provided in the medical charts using ICD-codes. Inconsistencies between coders, hospitals, countries, as well as the occasional lack of documentation in medical records lead to an information bias altering the indicators' accuracy [13–15]. This 'information bias' can produce perverse incentives, because under-coded data can generate falsely low indicator rates, while hospitals with a more complete chart documentation and more accurate coding would end up having higher measured rates of adverse events. The PSIs developed by the AHRQ are now used in a number of ways—including public reporting and pay-for-performance programs [2, 8]. The information bias just described, however, can encourage the maintenance of suboptimal data, because high PSI rates can pose both a reputational risk and financial risk to hospitals with higher quality data.

Combination of methodologies

To overcome the inaccuracy of these methodologies, a number of options could be considered. A first approach would be to avoid any reporting of quality and safety measures. However, this option is far from ideal, and some would even say unethical, considering the number of patients being harmed each year in health-care systems. A second approach would be to use only the voluntary reports of adverse events to develop corrective actions and to limit the preventable harms, without quantitative quality reporting. This option would limit the use of these reports to internal quality improvement efforts and would not be considered as a performance measure. A third approach, that we will elaborate fully in this paper, is to combine PSI measurement from administrative data with manual review of a only a small sample of medical records to provide accurate estimates of the true adverse event rates. We will demonstrate that it is possible to adjust measured PSI rates for validity, using the strengths of the manual review to limit the weaknesses of the PSIs. We now proceed to demonstrate why adjustment is so crucial, and how to do it.

Scenario 1: PSI #07 Central Venous Catheter-Related Blood Stream Infection Rate

We choose this PSI among the 27 PSIs developed by the AHRQ given it relative simplicity. To calculate this PSI #07, the algorithm selects as numerator all the hospitalizations with a secondary diagnosis of a catheter-related blood stream infection (CRBSI) in the administrative database among a denominator of all surgical and medical discharges, for patients aged 18 years and older. To introduce our validity-adjusted estimation, consider the following example using realistic numbers [14, 16]. Let us take an acute-care hospital with a volume of 100 000 stays. Without any adjustment, the algorithm finds 0.5% patients with an ICD-code related to a secondary diagnosis of CRBSI, resulting in 500 PSI #07 positive cases and 95 500 negative cases. However, we previously mentioned the potential inaccuracy of the PSI algorithms. Thus, it is necessary to adjust this PSI+ rate with other variables. This becomes possible if a hospital undertakes an explicit process of determining hospital-specific positive predictive value (PPV) and negative predictive value (NPV). This is possible to do using a manual review of a modest sample of medical charts. Indeed, a manual review of either all positive cases (for a rare PSI such as PSI#7) or a random sample of positive cases (e.g. $N = 200$–500, if the number of positives is high) to determine the PPV, and a random sample of negative cases ($N = 200$–500) to determine the NPV, would then give a hospital all information needed to determine validity-adjusted estimates of the true adverse event rate. If such a process were in place, and considering a hypothetical PPV of 85% and NPV of 99%, the steps to determine an estimated true PSI #07 rate would be as follows:

(i) Among 500 PSI+ cases, given known PPV = 85%, then we have:

$500 \times 0.85 = 425$ true positive (TP) and
$500 \times 0.15 = 75$ false positive (FP)

(ii) Among 99 500 PSI− cases, given known NPV = 99%, then we have:

$99\,500 \times 0.99 = 98\,505$ true negative (TN) and
$99\,500 \times 0.01 = 995$ false negative (FN)

(iii) Then, we can estimate the estimated true PSI rate as follows:

$(\text{TP} + \text{FN})/(\text{total number of patients}) = (425 + 995)/100\,000 = 1.42\%$

Importantly, this estimated true PSI rate of 1.42% is almost three times higher than the measured PSI+ rate of only 0.5%. To the extent that PPV and NPV measures are directly related to the accuracy of

Table 1 Three theoretical hospitals with variable quality of data and coding, and a presentation of how different estimated true adverse events rate are from the measured PSI#07 rate

Hospital	A	B	C
Data quality	High	Intermediate	Poor
Total number of patients	100 000	100 000	100 000
PSI rate (%)	*0.5*	*0.3*	**0.1**
Number of PSI+ cases	500	300	100
Number of PSI− cases	99 500	99 700	99 900
PPV (%)	85	70	60
NPV (%)	99	97	95
Estimated true PSI rate (%)	*1.42*	*3.20*	*5.06*

Table 2 Three theoretical hospitals with variable quality of data and coding, and a presentation of how different estimated true adverse events rate are from the measured global PSI rate for any adverse event

Hospital	A	B	C
Data quality	High	Intermediate	Poor
Total number of patients	100 000	100 000	100 000
PSI rate	**19%**	**15%**	**11%**
Number of PSI+ cases	19 000	15 000	11 000
Number of PSI− cases	81 000	85 000	89 000
PPV	55%	50%	45%
NPV	93%	90%	87%
Estimated true PSI rate	**16%**	**16%**	**17%**

coding, this adjustment of PSI rates based on validity potentially avoids the insidious effect of suboptimal data quality. Estimates of true PSI rates will be lower for hospitals with under-coding and higher for hospital with better coding. In Table 1, we provide a comparison of three hospitals with variable quality of coding.

With this concrete example, we compare the PSI #07 rates of three hospitals, with and without validity adjustment. The PSI+ rates presented without adjustment are the product of the quality of the data from which they are derived. The PSI algorithm finds lower rates for hospitals B and C because under-coded data generate lower PSI+ rates. However, in this example, hospital A collects data on PSI #07 more diligently, with higher PPV and NPV estimates, and these yield a higher crude rate of 0.5% for this PSI. After application of the validity-adjustment method, hospital C notably has an estimated true PSI #07 rate (5.06%) that is 50-fold higher than the measured PSI+ rate (0.1%). The relative ranking of these three hospitals is reversed, and most importantly, hospital A is not penalized for having better administrative data quality.

Scenario 2: Global PSI for Any Adverse Event

This second example presents the same approach for the global PSI for any adverse event developed by Southern *et al*. [17], using PPV and NPV estimates published elsewhere [18]. This global PSI is a non-rare event and, thus, presents a complementary example to the one presented above for rare events.

The PSI coding algorithm yields a higher rate of any adverse event for hospital A because better-coded data generate higher rates. After applying the validity-adjustment method, however, the estimated true PSI rates are close for the three hospital, increasing the rates of adverse events for hospitals B and C with intermediate or poor data quality, while actually *decreasing* the rate for the hospital with better-coded data (Table 2).

Key Data Ingredients Required for this Adjustment Method

We now proceed to outline the steps that a health-care system would need to follow to develop a system for measuring PSIs and then adjusting the PSIs to account for data validity. These include (i) a step of developing and running analytic algorithms for deriving PSI measurements; (ii) a step of undertaking a confined medical record review (i.e. a limited number of medical charts, to produce a measure of hospital-specific data validity) and (iii) a simple mathematical calculation to derive a validity-adjusted PSI rate.

Step 1: Algorithms for PSI The PSIs were developed by the AHRQ after a comprehensive literature review, analysis of ICD-9 codes, review by a clinician panel, implementation of risk adjustment and empirical analyses. The set of indicators had been secondarily adapted to the ICD-10, for use in international reporting initiatives, including one produced by the Organisation for Economic Co-operation and Development (OECD) [19]. Globally, the PSIs are derived by dividing a numerator created by searching for cases with any of a specified set of ICD-10 diagnosis codes (i.e. numerator) divided by the number of discharges at risk for the condition in question (i.e. denominator) (Box 1) [20]. Even though we chose the AHRQ PSIs to illustrate our methodological framework, the validity-adjustment method can also be used to adjust other PSIs using ICD-10 diagnosis codes published elsewhere [17].

Step 2: Medical chart review of trigger-positive cases (PSI+) to determine PPV and review of trigger-negative cases (PSI−) to determine NPV In order to determine estimates of hospital-specific PPV and NPV, a medical record review can be conducted using a standardized chart review protocol that directs reviewers to thoroughly review the entire chart (see detailed note in Supplemental Digital Content A). This approach to determining presence/absence of medical conditions triples the amount of time spent on chart review relative to traditional coding (10 to 30 minutes) and therefore creates a reference standard against which a PSI measurement (+ or −) can be compared [14, 21]. Importantly, the clinical reviewer must be blinded to whether they are reviewing the medical records of a PSI+ or PSI− case. The PPVs for each PSI will then be calculated as the proportion of PSI+ cases that are found to have the adverse event in question on detailed medical review. The NPVs will be calculated as the proportion of PSI− cases that are confirmed to not have the corresponding adverse event on detailed chart review. As mentioned earlier, such a medical record review could be conducted on all positive cases in a given hospital (for a relatively rare PSI such as PSI#7) or a random sample of positive cases ($N = 400$, sample size to achieve sufficiently precise estimate) to determine the PPV [22]. For determination of NPV, a random sample of negative cases ($N = 400$) would be selected for medical record review.

Step 3: Adjusted rate formula application Below, we present a simple Bayesian formula [23] to estimate true adverse rates (i.e. validity-adjusted event rates) by facility, while adjusting for imperfection in

Box 1 Algorithm for AHRQ PSI

AHRQ PSI:

PSI-02—death rate in low-mortality diagnosis-related groups (DRGs)
PSI-03—pressure ulcer rate
PSI-04—death rate among surgical inpatients with serious treatable conditions
PSI-05—retained surgical item or unretrieved device fragment count
PSI-06—iatrogenic pneumothorax rate
PSI-07—central venous catheter-related bloodstream infection rate
PSI-08—postoperative hip fracture rate
PSI-09—perioperative hemorrhage or hematoma rate
PSI-10—postoperative physiologic and metabolic derangement rate
PSI-11—postoperative respiratory failure rate
PSI-12—perioperative pulmonary embolism or deep vein thrombosis rate
PSI-13—postoperative sepsis rate
PSI-14—postoperative wound dehiscence rate
PSI-15—accidental puncture or laceration rate
PSI-16—transfusion reaction count
PSI-17—birth trauma rate—injury to neonate
PSI-18—obstetric trauma rate—vaginal delivery with instrument
PSI-19—obstetric trauma rate—vaginal delivery without instrument
PSI-90—patient safety for selected indicators

Example of PSI-13—postoperative sepsis rate

Denominator: elective surgical discharges, for patients aged 18 years and older, with any listed ICD-10-PCS procedure codes for an operating room procedure.

Numerator: discharges, among cases meeting the inclusion and exclusion rules for the denominator, with any of the following secondary ICD-10 diagnosis codes for sepsis:

A021: Salmonella sepsis
A227: Anthrax sepsis
A267: Erysipelothrix sepsis
A327: Listerial sepsis
A400: Sepsis due to Streptococcus, group A
A401: Sepsis due to Streptococcus, group B
A403: Sepsis due to Streptococcus pneumonia
A408: Other streptococcal sepsis
A409: Streptococcal sepsis, unspecified
A4101: Sepsis due to methicillin-susceptible Staphylococcus aureus
A4102: Sepsis due to methicillin-resistant Staphylococcus aureus
A411: Sepsis due to other specified Staphylococcus
A412: Sepsis due to unspecified Staphylococcus
A413: Sepsis due to Hemophilus influenza
A414: Sepsis due to anaerobes
A4150: Gram-negative sepsis, unspecified
A4151: Sepsis due to Escherichia coli
A4152: Sepsis due to Pseudomonas
A4153: Sepsis due to Serratia
A4159: Other Gram-negative sepsis
A4181: Sepsis due to Enterococcus
A4189: Other specified sepsis
A419: Sepsis, unspecified organism
A427: Actinomycotic sepsis
A5486: Gonococcal sepsis
B377: Candidal sepsis
R6520: Severe sepsis without septic shock
R6521: Severe sepsis with septic shock
T8110XA: Postprocedural shock unspecified, initial encounter
T8112XA: Postprocedural septic shock, initial encounter

one or both of PPV and NPV. The required input data are (i) the overall number of cases being screened with PSI (N_{total}); (ii) the number of PSI+ cases (N_{PSI+}); (iii) the number of PSI− cases (N_{PSI-}); (iv) PPV; and (v) NPV:

$$Estimated\ true\ PSI\ rate = \frac{[(N_{PSI+} \times PPV) + (N_{PSI-} \times (1 - NPV))]}{N_{total}}$$

Step 4: Validity adjusted rate report This rate adjustment calculation just shown can be applied to produce estimated true adverse event rates per hospital, for each AHRQ PSI or for new PSI reporting systems. The resulting 'validity-adjusted adverse event rates' can be descriptively compared to the crude PSI+ event rates by hospital, and they can also be juxtaposed with ancillary measures of facility performance (e.g. crude mortality rates, the hospital standardized mortality ratio [HSMR] [24] and volume of logged patient complaints by hospital).

Following these steps, participating hospitals will gain greater insight into their true adverse event rates, while also gaining knowledge on their hospital-specific data quality.

Discussion

There has long been discomfort around the use of PSIs derived from administrative data, because it is widely known that such data have imperfect validity and thus incomplete capture of adverse events arising in health care. Yet, alternative approaches to measuring adverse events also have well known problems. The limitations associated with alternative methods leave health-care systems with a conundrum—i.e. continue to measure adverse events using imperfect methods and imperfect data, or alternatively, just ignore adverse events and continue health-care operations in the dark, so to speak. We expect that most will reject the latter approach.

In this paper, we have demonstrated that there are actually ways to derive meaningful adverse event rate estimates, through the use of administrative databases. We have shown that PSI algorithms and measurements can be combined with a small-scale chart review and a simple Bayesian adjustment formula to determine a 'validity-adjusted adverse event rate'. These approaches take advantage of the notable strengths of administrative data. These data are produced routinely and comprehensively, and they permit the development of epidemiological indicators, the detailed description of health-care activities and, more generally, the assessment of health-care system performance. Above all, these databases can be analyzed using automated algorithms that can be applied to all hospital discharges, rather than just small subsets of medical charts.

Novel research is being conducted in the field of artificial intelligence, exploring automated artificial intelligence (AI) algorithms applied to electronic medical records. Even though machine learning and natural language processing are promising for the detection of adverse events in the medical notes, preliminary research sug-

gests that these methods will also not achieve a perfect validity [25]. Regardless of data type and analytic algorithms used, data will be imperfect, and consequently, indicators or trigger tools will be imperfect. In this context, validity-adjustment methodologies like the one shown here will continue to be needed.

Of central importance to the adjustment method that we have demonstrated here, the detection of the outcome of interest is strongly dependent on the quality of data reporting. The higher the quality of coding in a data system, the higher the number of adverse events that will be detected. Such a situation creates perverse incentives, unless there is an attempt to adjust PSI measurement rates to account for data quality and imperfect validity. Fortunately, the methodology that we demonstrate for PSIs creates a pathway for systems to simultaneously measure PSI rates and data validity, so that perverse incentives are avoided. It is highly undesirable to penalize a hospital that seeks to increase the quality of its data, while also producing higher PSI+ rates. The validity-adjustment method shown here would avoid that.

In relation to this last point, we underline that the quality of data at the hospital level may also in and of itself be an indicator of quality of care that is provided at that site. Indeed, such a notion would not surprise those who believe that good information is a foundation for good care. This underlines the fact that a hospital with a low measured PSI rate because of poor data quality may also have lower quality of care.

National or regional reporting systems could be structured around the general approach that we are proposing. First, unadjusted PSI+ can be generated using the administrative databases, just as is the case at present. The main change will be a chart review of a sample of PSI+ and PSI− cases, to estimate the PPV and NPV of the PSIs. For the AHRQ PSI system, a concrete application could be to adjust the 10 selected PSIs included in the CMS PSI-90 measure, because these 10 PSIs are a core metric in pay-for-performance (P4P) programs in the USA [1]. Implementation of a system where individual PSI rate estimates are adjusted in this manner will be more resource intensive, as the 10 PSIs in CMS PSI-90 will invoke a need to review a total of ∼4000 PSI+ and 4000 PSI− medical records. For the global PSI system, the review burden would be much lower, with only 400 PSI+ and 400 PSI− medical records needing review. To the extent that hospitals may have evolving data quality over time (perhaps even improving data quality over time, motivated somewhat by such PSI measurement systems), there is merit to considering such data validity assessments annually. Although this chart review will require allocation of resources (i.e. ∼$148 000 for the clinical reviewer salary and $20 000 for pulling the chart, in the CMS PSI-90 scenario, Supplemental Digital Content A), it will also provide high-quality data that could be used by existing hospital-based quality teams to develop corrective actions that can result in significant reductions in adverse events with resultant cost savings [26]. An adoption of this approach at a national or regional level could avoid unfair comparisons across hospitals and could also provide useful information for internal improvement for each hospital.

There are a few key points to clarify. First, it is important to not confuse the validity-adjustment procedure that we describe in this paper with clinical risk adjustment methods that should be applied to all comparative hospital outcome reports, where it is essential to adjust hospital-specific outcome rates for the baseline severity of illness of their patients. Such methods often include comparing observed outcome rates (observed deaths—O) with expected outcome rates derived from a prediction model based on the average severity of illness of patients treated at the hospital (expected deaths—E). Risk-adjusted outcome rates are then determined from the O/E ratio. The validity-adjustment method that we present, meanwhile, is an entirely different adjustment procedure that seeks to account for data validity within hospitals and across hospitals, when there is a desire to compare rates across facilities.

Second, our methodological proposal does not mention confidence intervals. This is not necessarily a problem, however, because the PSI estimates are not derived from a sample, but rather from a complete population. Assuming a binomial distribution for the PPV and NPV estimated by a chart review sample, it is still possible to obtain a range of values for the estimated true PSI rates using the PPV and NPV confidence intervals.

Third, no perfect method exists to measure adverse events in health care. Among all available methods, each with shortcomings, a manual review of medical records is commonly used. However, even though the methodology is widely believed to be more accurate than other systems (such as use of administrative data PSIs), the manual review remains imperfect, and the PPV and NPV estimates will be imperfect [11]. It is, therefore, important to pay attention to the methodology used for the manual review process, which requires some level of medical knowledge, clinical training and the use of a standardized chart review protocol (i.e. registered clinical nurse as reviewer).

Conclusion

The pursuit continues for better ways to measure and report on adverse events. The imperfect validity of administrative data is widely discussed and used by some as a black and white argument against their use. Here, however, we have demonstrated that there are ways to account for imperfect administrative data validity in PSI systems to derive estimates of true adverse event rates. We present potential steps that can be taken at a system level to create a framework for jurisdictional reporting of facility adverse event rates—a foundation for improvement in learning health systems.

Supplementary material

Supplementary material is available at *International Journal for Quality in Health Care* online.

Funding

This project benefitted from postdoctoral awards to Bastien Boussat from the University of Calgary's O'Brien Institute for Public Health and Cumming School of Medicine, the Grenoble Alpes Data Institute, which is supported by the French National Research Agency under the 'Investissements d'avenir' program (ANR-15-IDEX-02) and the Open Health Institute.

Contributorship

The authors confirm contribution to the paper as follows: study conception and design: WG; data simulation and analysis: BB; interpretation of results: all authors and draft manuscript preparation: BB, WG. All authors reviewed the results and approved the final version of the manuscript.

Ethics and other permissions

The PSI data presented are simulated. No new data were generated or analyzed.

Data availability

No new data were generated or analyzed in support of this review.

References

1. Centers for Medicare & Medicaid services. *Hospital-Acquired Condition Reduction Program Fiscal Year 2019 Fact Sheet*. 2019. https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Downloads/HAC-Reduction-Program-Fact-Sheet.pdf (14 February,2020, date last accessed).
2. French Ministry of Health. *Incitation Financière à l'Amélioration de la Qualité (IFAQ)*. https://www.has-sante.fr/portail/jcms/c_493937/fr/incitation-financiere-a-l-amelioration-de-la-qualite-ifaq-generalisation (14 February,2020, date last accessed).
3. National Health Service. *Transparency Data. Learning from Mistakes League*. https://www.gov.uk/government/publications/learning-from-mistakes-league (14 February,2020, date last accessed).
4. Mathes T, Pieper D, Morche J *et al.* Pay for performance for hospitals. *Cochrane Database Syst Rev* 2019;**7**:CD011156.
5. Ogundeji YK, Bland JM, Sheldon TA. The effectiveness of payment for performance in health care: a meta-analysis and exploration of variation in outcomes. *Health Policy* 2016;**120**:1141–50.
6. Rajaram R, Barnard C, Bilimoria KY. Concerns about using the Patient Safety Indicator-90 composite in pay-for-performance programs. *JAMA* 2015;**313**(9):897–8.
7. The Joint Commission. *Quality Check*. https://www.jointcommission.org/facts_about_quality_check_and_quality_reports/ (14 February,2020, date last accessed).
8. Medicare.gov. *Hospital Compare*. https://www.medicare.gov/hospitalcompare/about/what-is-HOS.html (14 February,2020, date last accessed).
9. Shetty P, Rivas R, Hristidis V. Correlating ratings of health insurance plans to their providers' attributes. *J Med Internet Res* 2016;**18**:e279.
10. Rafter N, Hickey A, Condell S *et al.* Adverse events in healthcare: learning from mistakes. *QJM* 2015;**108**:273–7.
11. Murff HJ, Patel VL, Hripcsak G *et al.* Detecting adverse events for patient safety research: a review of current methodologies. *J Biomed Inform* 2003;**36**:131–43.
12. Miller MR, Elixhauser A, Zhan C *et al.* Patient Safety Indicators: using administrative data to identify potential patient safety concerns. *Health Serv Res* 2001;**36**:110–32.
13. Kubasiak JC, Francescatti AB, Behal R *et al.* Patient Safety Indicators for judging hospital performance. *Am J Med Qual* 2017;**32**:129–33.
14. 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.
15. Ramanathan R, Leavell P, Stockslager G *et al.* Validity of Agency for Healthcare Research and Quality Patient Safety Indicators at an academic medical center. *Am Surg* 2013;**79**:578–82.
16. Cevasco M, Borzecki AM, O'Brien WJ *et al.* Validity of the AHRQ Patient Safety Indicator "central venous catheter-related bloodstream infections". *J Am Coll Surg* 2011;**212**:984–90.
17. 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.
18. McIsaac DI, Hamilton G, Abdulla K *et al.* Validation of new ICD-10-based Patient Safety Indicators for identification of in-hospital complications in surgical patients. *BMJ Qual Saf* 2020;**29**(3):209–216.
19. Drosler SE, Romano PS, Tancredi DJ *et al.* International comparability of Patient Safety Indicators in 15 OECD member countries: a methodological approach of adjustment by secondary diagnoses. *Health Serv Res* 2012;**47**:275–92.
20. Agency for Healthcare Research and Quality. *Guide to Patient Safety Indicators*. https://www.qualityindicators.ahrq.gov/downloads/modules/psi/v31/psi_guide_v31.pdf (14 February,2020, date last accessed).
21. Quan H, Sundararajan V, Halfon P *et al.* Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med Care* 2005;**43**:1130–9.
22. Hulley SB, Cummings SR, Browner WS *et al.* (eds). *Designing Clinical Research: An Epidemiologic Approach*. 4th edn. Appendix 6E. 2013, 81.Philadelphia: Lippincott, Williams & Wilkins.
23. Couris CM, Colin C, Rabilloud M *et al.* Method of correction to assess the number of hospitalized incident breast cancer cases based on claims databases. *J Clin Epidemiol* 2002;**55**:386–91.
24. Brien SE, Ghali WA. Public reporting of the hospital standardized mortality ratio (HSMR): implications for the Canadian approach to safety and quality in health care. *Open Med* 2008;**2**:e70–3.
25. Marella WM, Sparnon E, Finley E. Screening electronic health record-related patient safety reports using machine learning. *J Patient Saf* 2017;**13**:31–6.
26. Thanh NX, Baron T, Litvinchuk S. An economic evaluation of the National Surgical Quality Improvement Program (NSQIP) in Alberta, Canada. *Ann Surg* 2019;**269**:866–72.