"You can't improve until you measure": A Need Finding Study on Repurposed Clinical Indicators for Professional Learning

BERNARD BUCALON, The University of Sydney, Australia
CHRIS WILLIAMS, Adventist HealthCare Limited, Australia
JEANETTE CONLEY, Adventist HealthCare Limited, Australia
DAVID RANKIN, Cabrini Health, Australia
MARTIN VEYSEY, Royal Darwin Hospital, Australia
TIM SHAW, The University of Sydney, Australia
JUDY KAY, The University of Sydney, Australia

Hospitals generate large amounts of administrative data from healthcare delivery. Much of this data is for summative quality and safety (Q&S) reporting rather than formative professional learning. Little is known about what clinicians want to learn about their practice from this data and how they want to see that information. To address this, we designed a need finding study, run at a large metropolitan hospital. This paper describes the study design, based on semi-structured interviews which used four static visualisations prototypes. Our participants, nine specialists valued the re-purposing of clinical indicators (CIs) for professional learning. They particularly valued peer comparison and trends for benchmarking and highlighted the need for risk-adjusted results that are carefully selected to ensure fair and meaningful comparisons. Our contributions are: (1) insights and needs from participant free comments, (2) comments around our four visualisation prototypes, and (3) design recommendations when presenting CIs for professional learning.

CCS Concepts: • **Human-centered computing** → *User studies*; *Empirical studies in visualization.*

Additional Key Words and Phrases: clinical practice analytics, clinical indicators, data visualisation, professional learning, reflective practice, continuing professional development

ACM Reference Format:
Bernard Bucalon, Chris Williams, Jeanette Conley, David Rankin, Martin Veysey, Tim Shaw, and Judy Kay. 2018. "You can't improve until you measure": A Need Finding Study on Repurposed Clinical Indicators for Professional Learning. In . ACM, New York, NY, USA, 12 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Hospitals routinely collect large amounts of administrative data through the day-to-day operations of healthcare delivery. There is potential to re-purpose this data to support professional development, providing a new and valuable foundation for continuous healthcare improvement [17]. It could support reflective practice that also meets new Continuing Professional Development (CPD) requirements related to reviewing performance and measuring patient outcomes [31]. This is quite different from well established processes of quality and safety (Q&S) reporting that is driven by hospital-level performance goals. Reflective practice for CPD is driven by individual professional goals and poses new system design and user interface challenges [40].

Our work emerged from needs identified at a private hospital in Sydney, Australia. They wanted their individual clinicians and specialty groups to have access to a suitable form of clinical indicators (CI) based on administrative data, such as length of stay, 28-day readmissions, and complications. They had centralised the reporting of routinely collected *process* and *outcome* performance indicators [14]. *Microsoft PowerBI*, a self-service analytics platform, had already enabled medical administrators and clinical informaticians to search, filter, and visualise hospital-wide performance indicators. This meant they no longer relied on ICT staff to write bespoke queries and generate reports about this data. The hospital leadership saw the potential learning benefits of making this information available to clinicians. The work is part of a larger research program exploring clinical practice analytics [25].

One barrier in creating such interfaces is that little is currently known about what clinicians want to learn about their own practice from such data [5]. Beyond this, it is important to discover how to present the information in ways that are understandable and useful [19]. This led us to formulate the following research questions:

(1) What do clinicians want to know about their practice from routinely collected CIs?
(2) What are the key design considerations when presenting CI data for professional learning?

We contribute to the emerging literature on re-purposed hospital data for learning and development [37]. While studies have evaluated clinical dashboards for decision-making in patient management, hospital administration, and monitoring adherence to Q&S standards [15, 27]; our need finding study explored what clinicians want to know about their practice from CI data for professional learning and provides foundations for designing interfaces that they can understand and relate to their own professional development.

2 RELATED WORK

There are several challenges in repurposing routinely collected CIs for professional learning. Some follow from the fact that Electronic Medical Records (EMR) and patient administration systems were designed primarily for patient management, billing, and to monitor care outcomes [35, 44].Challenges also come from the nature of the data; it can be incomplete and have errors introduced at data entry and in coding diagnoses and procedures [20]. This is particularly common when the person entering the data is unaware of ways the data may be used.

Professional learning is largely informal and experiential [9, 16]. So CI data is an incomplete representation of individual performance [40]. Clinicians may practice both as individuals or in a multi-disciplinary team (MDT), often making it difficult to interpret outcome data in terms of individual performance [22]. Administrative process indicators can be used across specialties, but CI selection is still critical. What teams want to know about their practice varies between medical and surgical specialties [30]. A fourth challenge is fair comparison of peer performance. This is difficult as each clinician treats different patient populations. Whilst imperfect, risk-adjustment is used to account for the complexity in patient demographics, co-morbidities, and the surgical procedures performed [12, 38].

Beyond these data challenges, another key problem that we aim to tackle is how to design relevant and useful presentations of the data. This involves two challenging design problems: (1) identifying suitable ways to present complex information and (2) communicating the varying levels of uncertainty across the data. HCI research methods have been used to better design tools to present complex electronic health data. For example, Seals et al. [39] used interviews and a think-aloud protocol to understand the needs of clinicians when visualising wearable sensor data. Jo et al. [26] conducted need finding and feedback interviews with medical practitioners to iteratively design a clinical decision support tool for prescribing psychiatric drugs. Backonja et al. [3] conducted a user-centred study with healthcare professionals to design data visualisations for personalised care for cancer patients. A key challenge for design is that people find uncertainty difficult to understand [24, 46]. Even error bars in bar charts can pose problems [10]. Further, evaluating data visualisations is difficult [28, 36]. While there have been studies on design and evaluation of data visualisations in clinical settings, few have explored the presentation of repurposed CIs for professional learning and reflective practice [8].

3 STUDY DESIGN

To answer our research questions, we designed an interview-based need finding study [29]. This began with broad questions about the participants' background and perceived needs and wishes. Then we presented a set of four prototype interfaces and asked participants to respond to these. We now describe the details of the stages and the rationale.

3.1 Recruitment and Interview protocol

The study was conducted at a large metropolitan private hospital in Sydney, Australia. We recruited a purposeful sample of clinicians through an email invitation distributed to staff specialists (also referred to as "consultants") [43]. We also employed snowball recruitment [11]. We used a semi-structured interview approach. Permission to conduct this study was received from the Adventist HealthCare Limited (AHCL) Human Research Ethics Committee (HREC) (2020-035).

To start the interview, (1) We confirmed their current specialty, years of experience, and how they worked as part of a larger team. (2) We then asked interviewees to define reflective practice in their own words. This enabled us to better understand their perceptions about the process of reflection. (3) We asked participants to imagine they had a magic wand and to describe an ideal reflective practice dashboard. (4) We then asked them to identify the clinical indicators relevant to their practice; to describe how these CIs should be presented; and describe specific functionality ("features") they felt would be useful. We then asked interviewees to confirm the two most important clinical indicators for their practice, to prioritise the CIs already mentioned. (5) After this general, broad elicitation, we asked participants to think-aloud [29] as they make use of four quite different data visualisation design prototypes. This approach, modelled on the approach of [42], aimed to gain rich comments and critiques of diverse ways to present clinical indicator information. Participants were asked to describe what that liked and disliked about each visual design. (6) Finally, we gave interviewees an opportunity to change or add to their initial responses about their ideal reflective practice dashboard.

3.2 Design of the visualisations

We wanted a small number of visualisations so that participants could study them in a one hour session. We designed these for our target audience of clinicians. We took account of research indicating what causes difficulties in reading visualisations [19]. We also considered how clinicians make heavy use of visual data and graphs such as histograms. Nonetheless, we were careful not to assume that any depiction of uncertainty would be easy to understand [18, 23].

We sourced real world clinical indicator data from two research partner private hospitals, based on routinely extracted Hospital Casemix Protocol (HCP) data. HCP is a legislated data collection of all private health insurance funded admissions [41]. The HCP dataset does not include information on the admitting or operating medical practitioner, so we merged that information to produce authentic doctor-specific performance data.

We then reviewed examples of quarterly reports generated from the HCP data used in practice during Orthopaedic team meetings. The authors with a clinical background, staff from the hospital and experts in clinical informatics, identified the following four standard CIs to use in the prototype: *length of stay*, *28-day readmissions*, *hospital acquired complications (HAC)*, and *deaths*, because they are important CIs that the participants would be familiar with. To identify the display formats to evaluate, we drew on a scoping review of interfaces based on routinely collected CI data to support reflection on practice [8]. Based on the most common forms that appeared in the work, we designed the prototype around the four display formats: *bar chart*, *box-and-whisker plot*, *scatter plot*, and *table*.

The first author mocked up the static prototypes in *Figma*. To ensure anonymity of clinicians in the HCP dataset, we excluded any data that may be identifiable due to low case loads. The first author then met with the co-authors to incorporate their feedback into the prototype. We now briefly describe the prototypes - all appear in Appendix A.1.

Bar chart. (Appendix A.1 Figure 1). We wanted a data visualisation of individual clinician performance, comparing two common clinical indicators. The bar chart shows six months of data on the top Diagnosis-Related Groups (DRG) by Number of Patient Episodes. For each DRG, it shows the Average Length of Stay (LOS). Average LOS is commonly linked to Episodes to account for different clinician case loads.

Box-and-whisker plot. (Appendix A.1 Figure 2). This compares an individual surgeon's readmission rates against their peers to highlight two things: first, how close their readmission rate was to the median, and second, highlight any outliers. For each calendar quarter, the box plot counts the number of patients re-admitted within 28 days of discharge and plots the distribution with a box-and-whisker plot. Unplanned hospital readmission rate is considered as a performance indicator to measure quality of care [45].

Scatter plot. (Appendix A.1 Figure 3). We wanted to cluster clinicians based on the complication rates to highlight similar clinicians in a team. For the past six months, the chart plots all Respiratory Medicine physicians by their Complication Rate and their Number of Patient Episodes. Patient hospital acquired complications are presented as a rate (e.g. of 10,000 episodes) to adjust for varying clinician case loads.

Table. (Appendix A.1 Figure 4). This was benchmark data for one hospital against comparable hospitals at state and national level. From 2014 to 2019, the table shows a count the number of patient Deaths that had Delays in Surgical Diagnosis. Deaths with delayed surgical diagnosis are common CIs of quality of care [34].

3.3 Data analysis

The first author manually transcribed the audio recordings, removing any identifiable information. They then open-coded each transcript by highlighting quotes that were relevant or potentially interesting, then assigned descriptive codes [7]. Codes were then grouped into relevant themes and sub-themes. The first and last authors then discussed the disagreements over multiple meetings, to refine the themes until there was consensus.

4 FINDINGS

Table 1 shows the characteristics of the nine clinicians we interviewed (2 female). The nine clinicians are from different specialty groups (5 surgical). Their experience ranged between 1.5 to 40.0 years (mean 18.2). Only two (of 9) interviews were conducted in-person at the hospital; the others were conducted online. The participants provided rich feedback on interpreting common CIs for professional learning. We present their experiences using CIs to change practice, considerations for data interpretation and presentation. Then we detail feedback on each data visualisation design.

4.1 RQ1

Reflective practice. We asked participants to define reflective practice in their own words. They described the process as understanding the gap between what they think (in their minds) and what they know (the data) [P01]. It occurs

Table 1. Characteristics of participants, N=9

ID	Specialty and sub-specialties	Medical / Surgical	Sex	Years experience	Interview
P01	Interventional Pain and Palliative Medicine	Medical	M	20.0	In-person
P02	Cardiology	Medical	M	37.0	Online
P03	Colorectal and General Surgery	Surgical	M	1.5	Online
P04	Orthopedic Surgery	Surgical	F	22.0	Online
P05	Breast, Endocrine and General Surgery	Surgical	M	7.0	Online
P06	Upper GI, Bariatric and Laparoscopic Surgery	Surgery	M	10.0	In-person
P07	Surgical Oncology	Surgical	M	25.0	Online
P08*	General Practice, Rehabilitation Medicine	Medical	M	1.5	Online
P09	Anesthesia	Medical	F	40.0	Online

* P08 had to leave and did not provide feedback on the designs.

constantly and iteratively, by comparing their expectations of outcomes before and after [P03, P04]. One said they had no strict definition, it was simply "thinking about your own practice" [P08].

> *"And so that's very powerful for changing behaviours. And if someone sees that there's a bunch of people at a hundred percent and they are down at 30%, one would hope that would, might encourage them to change their behaviour."* **[P07]**

In general, they were interested in what their colleagues think. Some described this as helping identify what went wrong and how things could be better [P07, P09]. One participant said measurement was needed for improvement, that without the data they "felt lost" [P05]. Another said they keep a list of interesting cases to periodically review [P08].

Clinical indicators. We asked participants to identify clinical indicators that were relevant to their individual practice. We categorise the CIs as *process*, *outcome*, and *patient* experience indicators. The process indicators included: length of stay [P02, P05], readmissions [P04], complications [P05, P06], case load/throughput [P01, P06], completed discharge summaries [P06], and time spent in recovery [P09]. The outcome indicators include: pain score [P09], weight loss [P06], morbidity [P07], and death [P02]. Most participants highlighted the importance of patient reported outcomes and experience (PROMs and PREMs) in judging individual and team performance [P01, P02, P04, P06-P09]. Capturing patient satisfaction and experience through surveys were key for most.

> *"One of the most important things in a private practice is patient satisfaction. How quickly they get to see a patient? How quickly did they communicate with the referring doctor? Did the discharge letter get out fast enough? How well are they communicating with the patient, with the families, with the referring doctor, with their colleagues, et cetera. There's so many different elements."* **[P01]**

Interpretation. Participants highlighted the need for benchmarked data to compare individual and team indicators with like-for-like units across the state, country, and internationally [P04, P05]. Risk-adjustment of the clinical indicators was mentioned to account for the variance in patient co-morbidities and complexity [P07]. Participants also noted the risk of misinterpreting broad, process indicators such as length of stay:

> *"Now I think this is a really bad indicator (LOS) because of the fact that the patients in recovery for longer doesn't necessarily mean something went wrong. It may mean that they have some reason that you want to keep them under close observation for longer."* **[P09]**

Presentation. We asked participants to comment on how best to present re-purposed CIs to support reflection on their practice. They wanted to see their CIs displayed as trends, to compare themselves over last week, month, and year [P05]. They were interested in trends in patient selection and seasonal peak and troughs [P01, P05]. They expected de-identified data to maintain confidentiality, either through summarised views or peer-to-peer comparison data [P07].

> *"For example, you want to know which direction you're going, to be able to predict the future. If you see one particular type of patient – you would be more interested in your practice – you need to be prepared to respond to that demand in the future. So I'm interested in trends basically."* **[P05]**

4.2 RQ2

Table 2 summarises feedback on the four data visualisation designs. The leftmost sub-table shows the overall reactions. The next block shows if the participants interpreted the information correctly. The other two sub-tables show counts of the questions and suggestions from participants.

Table 2. Summary of design feedback: In Reaction, Pos means positive, Neg means negative and Net means Neither,

	Reaction				Interpretation				Questions				Suggestions			
PID	**D1**	**D2**	**D3**	**D4**	**D1**	**D2**	**D3**	**D4**	**D1**	**D2**	**D3**	**D4**	**D1**	**D2**	**D3**	**D4**
P01	Net	Net	Neg	Pos	N	Y	Y	Y	1							
P02	Neg	Neg	Neg	Neg	N	N	Y	Y	1	3	1	1	1			1
P03	Net	Net	Neg	Net	Y	Y	Y	Y		1		1				
P04	Neg	Neg	Neg	Neg	Y	Y	N	Y	2		1			2		3
P05	Pos	Pos	Pos	Neg	Y	Y	Y	Y							3	1
P06	Net	Neg	Neg	Pos	Y	Y	N	Y		1	2					1
P07	Neg	Pos	Pos	Net	N	Y	Y	Y							2	2
P09	Neg	Neg	Neg	Net	Y	N	N	Y	1	3	2	2				

Note: **D1**=Bar chart of Average Length of Stay (LOS). **D2**=Box plot of 28-day Readmissions. **D3**=Scatter plot of Hospital Acquired Complication (HAC) Rate. **D4**=Table of Deaths with Delay in Surgical Diagnosis. * P08 has been omitted from the table.

D1. Bar chart of Average Length of Stay. Four participants made negative comments, for example, saying they were "confused" or it was "hard to interpret" [P02]. One was unsure if the chart was individual or for a team of clinicians [P01]. One was unable to interpret LOS and episode count on the same chart [P07]. Five could interpret it correctly. The main problem was interpreting two series on one chart: Number of Episodes **(B)** on the bottom x-axis and Average Length of Stay **(C)** on the top x-axis, as in this comment from **P04**:

> *"I can't actually, I don't know, there's a different, it's different across the top and on the bottom, which doesn't make sense to me."*

D2. Box plot of 28-day Readmissions. The two positive participants said it was a good way to present 28-day readmissions and to show "where you are in the data" [P05, P07]. Four participants were negative, including the two who could not interpret it. Comments described it as complicated and hard to interpret [P02, P04, P06, P09], for example:

> *"I don't know whether I would have chosen to compare something like that over time like that. I dunno if they introduced some new things. Some new protocol or some procedure. I dunno know what you're wanting to get out of that."* **[P09]**

P04 suggested improving it by breaking down the data by the diagnosis (ICD-10 code) [1] or procedure (MBS item number) [2], to make it easier to see what factors contributed to the readmission.

D3. Scatter plot of Hospital Acquired Complication (HAC) Rate. Six participants were negative and three of these participants could not interpret it. They found it was confusing and did not realise the data was about a team of Respiratory Physicians, rather than an individual [P04, P09]. They had trouble with the HAC figure **(G)**. One did not notice the complication rate out of 10,000 episodes [P06]. P09 did not notice that the "minimum 25 episodes" line **(H)**, excluding clinicians with case loads too small to compare fairly. **P07** was positive about it and could interpret it said:

> *"Anyway, I just think at face value it's reasonable to interpret it, but you can't read too much into it because you don't know the nature of in fact what different people are doing and the patients they're working on."*

One participant suggested the Hospital Acquired Complications (HAC) need to be categorised to better make sense of this data. Without the HACs broken down, for example, one cannot know whether the complication was an "infection or pneumonia" [P05].

D4. Table of Deaths with Delay in Surgical Diagnosis. This had all participants correctly interpret comparisons of like-hospitals at state and national level. Three were negative, while P04 and P05 felt it would be more useful to break down the data by surgical specialty or procedure. P06 suggested to display statistical significance. Another suggested to add a bar chart to compliment the raw data in the table. **P03** had concerns about the definition of "like-hospitals":

> *"But when comparing to government and national hospitals the question is, is their auditing as good as our auditing? And are we picking up more while there a lot of things may be missed? Because they may not have funding to have the coders and the necessary data collectors."*

5 DISCUSSION, RECOMMENDATIONS, AND FUTURE WORK

Our study identified what clinicians want to learn about their practice from re-purposed CI data and their comments gave valuable insights for designing future interfaces. We now summarise key insights and how they can inform design.

Valuing relevant re-purposed CI data for reflective practice. All participant descriptions of reflective practice aligned with the literature. They described it as deliberate thinking about experiences to learn from mistakes, identifying areas for improvement and actions to improve future outcomes [16]. All supported the goal of re-purposing routinely collected administrative data for learning. However, they emphasised the importance of selecting clinical indicators that are relevant to their unique scope of practice [20].

- **R1: Explain how CIs are collected and calculated**: Clarify definitions of CIs as they may differ between jurisdictions. E.g. Late Discharge: Separation time of patient episode after 10am; Same Day Patient: admitted and discharged on the same date.

Key factors interact. Participants were cognizant that one data visualisation could be interpreted different ways, for example commenting: "data itself does not tell to the whole story". This was best illustrated in an example from the anaesthetist [P09]. They described the tension between high level administrative KPIs vs. lower level patient centred indicators. A surgeon with slightly higher average Length of Stay than their peer cohort may also have better patient satisfaction scores than peers. This matches findings in Brasel et al. [6], where LOS should not be used as a single outcome indicator due to the influence of non-clinical factors, such as socioeconomic status and social support networks.

- **R2: Avoid single process and outcome indicators**: Combine CIs to present a more complete picture. Select CIs individual clinicians and teams can improve, rather than high level administrative priorities. E.g. LOS with Patient Satisfaction; Complication Rate with Number of Episodes.

Trends and fair comparisons. On data interpretation, they expressed concerns that administrators may use this data punitively for performance management, rather than for formative professional development [40]. Most wanted to see trends, both comparing their past self over multiple time periods, and also against local and international benchmarks [32]. Most participants gave valuable criticisms of the four data visualisations. They cautioned the use of comparators based on "blunt" indicators from administrative data; they stressed that clinicians should only be compared with "like-for-like" peers. One approach to account for patient case complexity is to risk-adjust the underlying data [12].

- **R3: Breakdown broad CIs**: Use other dimensions to filter broad CIs down into smaller usable groups. E.g. By MBS item number (procedure); Hospital Acquired Complication (HAC) category; urgency of admission.
- **R4: Consider risk-adjusting CIs**: Where appropriate, adjust CIs by patient demographics, co-morbidities, and procedure complexity. E.g. Risk-Adjusted LOS.

Peer comparison and benchmarking. All participants mentioned these. Access to data about their peers was seen as valuable in identifying gaps in individual performance and promoting practice change. This is in line with the different use of such data for Audit and Feedback (A&F), a quality improvement process which focuses on actual and desired performance [13]. Gude et al. [21], investigated the best practices in selecting A&F performance comparators. Of the 146 trials evaluated, 60% used benchmarks, while 13% used a combination of trends or explicit targets. They recommend providing tailored comparisons rather than benchmarking everyone against the mean.

- **R5: Use different comparators**: Provide options for peer comparison. E.g. Individual vs past performance, Individual vs. group, Individual vs. peer.
- **R6: Signpost individual vs. grouped data**: Indicate whether individual or grouped data is displayed. Use a distinctive and consistent style. E.g. "Your average LOS", "Group average LOS".

Limitations. Key limitaions are: The study was conducted at a single hospital site; Prototypes were shown in the same order in each interview; One author transcribed and coded interviews.

6 CONCLUSIONS AND FUTURE PLANS

We present a need finding study to understand what clinicians want to learn about their practice from CI data. We captured rich insights for presenting this data, by crafting four static but diverse visualisations and asking participants to interpret the data. Participants saw value in repurposing this data for professional learning. The findings highlight considerations around CI selection, data interpretation, and presentation of benchmarking and trend data for peer comparison. The current study will shape the design of a high fidelity dashboard prototype. This will be evaluated in a long-term multi-site field study involving both team meetings and individual use by diverse specialty groups such as Neurology, Urology and a Cancer MDT (Medical Oncology, Radiation Oncology, Palliative Medicine specialties). The results from the field study will help inform the design of future tools to support clinician reflective practice for CPD.

ACKNOWLEDGMENTS

The research described in this paper was supported by Digital Health CRC Limited ("DHCRC"), funded under the Commonwealth's Cooperative Research Centres (CRC) Program. Author BB holds a higher degree of research scholarship funded by DHCRC. We thank Anna Janssen for assistance in the interviews. We thank the participants for their time and sharing their experiences.

REFERENCES

[1] AIHW. 2022. Codes and classifications. https://www.aihw.gov.au/suicide-self-harm-monitoring/data/technical-notes/codes-and-classifications

[2] AIHW. 2022. Medicare Benefits Schedule (MBS) data collection. https://www.aihw.gov.au/about-our-data/our-data-collections/medicare-benefits-schedule-mbs

[3] Uba Backonja, Sarah C Haynes, and Katherine K Kim. 2018. Data Visualizations to Support Health Practitioners' Provision of Personalized Care for Patients With Cancer and Multiple Chronic Conditions: User-Centered Design Study. *JMIR Human Factors* 5, 4 (Oct. 2018), e11826. https://doi.org/10.2196/11826

[4] Tim Badgery-Parker, Sallie-Anne Pearson, Susan Dunn, and Adam G. Elshaug. 2019. Measuring Hospital-Acquired Complications Associated With Low-Value Care. *JAMA Internal Medicine* 179, 4 (April 2019), 499. https://doi.org/10.1001/jamainternmed.2018.7464

[5] Elisa Bindels, Christel Verberg, Albert Scherpbier, Sylvia Heeneman, and Kiki Lombarts. 2018. Reflection revisited: how physicians conceptualize and experience reflection in professional practice – a qualitative study. *BMC Medical Education* 18, 1 (Dec. 2018), 105. https://doi.org/10.1186/s12909-018-1218-y

[6] Karen J. Brasel, Hyun J. Lim, Ram Nirula, and John A. Weigelt. 2007. Length of stay: An appropriate quality measure? *Archives of Surgery* 142, 5 (May 2007), 461–466. https://doi.org/10.1001/archsurg.142.5.461 tex.eprint: https://jamanetwork.com/journals/jamasurgery/articlepdf/400154/sws70002_461_466.pdf.

[7] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (Jan. 2006), 77–101. https://doi.org/10.1191/1478088706qp063oa

[8] Bernard Bucalon, Tim Shaw, Kerri Brown, and Judy Kay. 2022. State-of-the-art Dashboards on Clinical Indicator Data to Support Reflection on Practice: Scoping Review. *JMIR Medical Informatics* 10, 2 (Feb. 2022), e32695. https://doi.org/10.2196/32695

[9] Mark Cole. 2000. Learning through reflective practice: a professional approach to effective continuing professional development among healthcare professionals. *Research in Post-Compulsory Education* 5, 1 (March 2000), 23–38. https://doi.org/10.1080/13596740000200067

[10] Michael Correll and Michael Gleicher. 2014. Error Bars Considered Harmful: Exploring Alternate Encodings for Mean and Error. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (Dec. 2014), 2142–2151. https://doi.org/10.1109/TVCG.2014.2346298

[11] John W. Creswell. 2018. Qualitative inquiry & research design : choosing among five approaches. In *Qualitative inquiry & research design : choosing among five approaches* (fourth edition. ed.). SAGE Publications, Incorporated, Los Angeles ;.

[12] Susan I. DesHarnais, Thane M. Forthman, Judith M. Homa-Lowry, and Dwight L. Wooster. 2000. Risk-Adjusted Clinical Quality Indicators: Indices for Measuring and Monitoring Rates of Mortality, Complications, and Readmissions. *Quality Management in Health Care* 9, 1 (2000), 14–22. https://doi.org/10.1097/00019514-200009010-00003

[13] Laura Desveaux, Noah Michael Ivers, Kim Devotta, Noor Ramji, Karen Weyman, and Tara Kiran. 2021. Unpacking the intention to action gap: a qualitative study understanding how physicians engage with audit and feedback. *Implementation Science* 16, 1 (Dec. 2021), 19. https://doi.org/10.1186/s13012-021-01088-1

[14] Avedis Donabedian. 2005. Evaluating the Quality of Medical Care. *The Milbank Quarterly* 83, 4 (2005), 691–729.

[15] Dawn Dowding, Rebecca Randell, Peter Gardner, Geraldine Fitzpatrick, Patricia Dykes, Jesus Favela, Susan Hamer, Zac Whitewood-Moores, Nicholas Hardiker, Elizabeth Borycki, and Leanne Currie. 2015. Dashboards for improving patient care: Review of the literature. *International Journal of Medical Informatics* 84, 2 (Feb. 2015), 87–100. https://doi.org/10.1016/j.ijmedinf.2014.10.001

[16] A. DunnGalvin, Jeffrey B. Cooper, G. Shorten, and Richard H. Blum. 2019. Applied reflective practice in medicine and anaesthesiology. *British Journal of Anaesthesia* 122, 5 (May 2019), 536–541. https://doi.org/10.1016/j.bja.2019.02.006

[17] Joanne Enticott, Sandra Braaf, Alison Johnson, Angela Jones, and Helena J. Teede. 2020. Leaders' perspectives on learning health systems: a qualitative study. *BMC Health Services Research* 20, 1 (Dec. 2020), 1087. https://doi.org/10.1186/s12913-020-05924-w

[18] Michael Fernandes, Logan Walls, Sean Munson, Jessica Hullman, and Matthew Kay. 2018. Uncertainty Displays Using Quantile Dotplots or CDFs Improve Transit Decision-Making. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Montreal QC Canada, 1–12. https://doi.org/10.1145/3173574.3173718

[19] Steven L. Franconeri, Lace M. Padilla, Priti Shah, Jeffrey M. Zacks, and Jessica Hullman. 2021. The Science of Visual Data Communication: What Works. *Psychological Science in the Public Interest* 22, 3 (Dec. 2021), 110–161. https://doi.org/10.1177/15291006211051956

[20] Tim Freeman. 2002. Using performance indicators to improve health care quality in the public sector: a review of the literature. *Health Services Management Research* 15, 2 (May 2002), 126–137. https://doi.org/10.1258/0951484021912897

[21] Wouter T Gude, Benjamin Brown, Jamie C Brehaut, Zach Landis-Lewis, Christopher J Armitage, Nicolette F de Keizer, and Niels Peek. 2019. Clinical performance comparators in audit and feedback: a review of theory and evidence. *Implementation Science* 14, 1 (2019), 14.

[22] Carrie A. Herzke, Henry J Michtalik, Nowella Durkin, Joseph Finkelstein, Amy Deutschendorf, Jason Miller, Curtis Leung, and Daniel J Brotman. 2018. A Method for Attributing Patient-Level Metrics to Rotating Providers in an Inpatient Setting. *Journal of Hospital Medicine* 13, 7 (July 2018), 470–475. https://doi.org/10.12788/jhm.2897

[23] Jake M. Hofman, Daniel G. Goldstein, and Jessica Hullman. 2020. How Visualizing Inferential Uncertainty Can Mislead Readers About Treatment Effects in Scientific Results. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–12. https://doi.org/10.1145/3313831.3376454

[24] Jessica Hullman, Xiaoli Qiao, Michael Correll, Alex Kale, and Matthew Kay. 2019. In Pursuit of Error: A Survey of Uncertainty Visualization Evaluation. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (Jan. 2019), 903–913. https://doi.org/10.1109/TVCG.2018.2864889

[25] Anna Janssen, Stella Talic, Dragan Gasevic, Judy Kay, and Tim Shaw. 2021. Exploring the Intersection Between Health Professionals' Learning and eHealth Data: Protocol for a Comprehensive Research Program in Practice Analytics in Health Care. *JMIR Research Protocols* 10, 12 (Dec. 2021), e27984. https://doi.org/10.2196/27984

[26] Eunkyung Jo, Myeonghan Ryu, Georgia Kenderova, Samuel So, Bryan Shapiro, Alexandra Papoutsaki, and Daniel A. Epstein. 2022. Designing Flexible Longitudinal Regimens: Supporting Clinician Planning for Discontinuation of Psychiatric Drugs. In *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–17. https://doi.org/10.1145/3491102.3502206

[27] Saif Sherif Khairat, Aniesha Dukkipati, Heather Alico Lauria, Thomas Bice, Debbie Travers, and Shannon S Carson. 2018. The Impact of Visualization Dashboards on Quality of Care and Clinician Satisfaction: Integrative Literature Review. *JMIR Human Factors* 5, 2 (May 2018), e22. https://doi.org/10.2196/humanfactors.9328

[28] H. Lam, E. Bertini, P. Isenberg, C. Plaisant, and S. Carpendale. 2012. Empirical Studies in Information Visualization: Seven Scenarios. *IEEE Transactions on Visualization and Computer Graphics* 18, 9 (Sept. 2012), 1520–1536. https://doi.org/10.1109/TVCG.2011.279

[29] Jonathan Lazar. 2017. *Research methods in human-computer interaction* (second edition. ed.). Morgan Kaufmann Publishers, an imprint of Elsevier, Cambridge, MA. Publication Title: Research methods in human-computer interaction.

[30] J. Mainz. 2003. Defining and classifying clinical indicators for quality improvement. *International Journal for Quality in Health Care* 15, 6 (Dec. 2003), 523–530. https://doi.org/10.1093/intqhc/mzg081

[31] MBA. 2017. *Expert Advisory Group on Revalidation Final Report.* Technical Report. Medical Board of Australia. http://www.medicalboard.gov.au/documents/default.aspx?record=WD17%2F24295&dbid=AP&chksum=Txmn8C7v%2BC53Wjsz3sXn2w%3D%3D

[32] Craig McFadyen, Sara Lankshear, Dimitrios Divaris, Mark Berry, Amber Hunter, John Srigley, and Jonathan Irish. 2015. Physician level reporting of surgical and pathology performance indicators: a regional study to assess feasibility and impact on quality. *Canadian Journal of Surgery* 58, 1 (Feb. 2015), 31–40. https://doi.org/10.1503/cjs.004314

[33] Natasa Mihailovic, Sanja Kocic, and Mihajlo Jakovljevic. 2016. Review of Diagnosis-Related Group-Based Financing of Hospital Care. *Health Services Research and Managerial Epidemiology* 3 (Jan. 2016), 233339281664789. https://doi.org/10.1177/2333392816647892

[34] J B North, F J Blackford, D Wall, J Allen, S Faint, R S Ware, and T Rey-Conde. 2013. Analysis of the causes and effects of delay before diagnosis using surgical mortality data. *British Journal of Surgery* 100, 3 (Jan. 2013), 419–425. https://doi.org/10.1002/bjs.8986

[35] Sun Young Park and Yunan Chen. 2012. Adaptation as design: learning from an EMR deployment study. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Austin Texas USA, 2097–2106. https://doi.org/10.1145/2207676.2208361

[36] Catherine Plaisant. 2004. The challenge of information visualization evaluation. In *Proceedings of the working conference on Advanced visual interfaces - AVI '04*. ACM Press, Gallipoli, Italy, 109. https://doi.org/10.1145/989863.989880

[37] Martin V. Pusic, Robert J. Birnbaum, Brent Thoma, Stanley J. Hamstra, Rodrigo B. Cavalcanti, Eric J. Warm, Anna Janssen, and Tim Shaw. 2022. Frameworks for Integrating Learning Analytics With the Electronic Health Record. *Journal of Continuing Education in the Health Professions* Publish Ahead of Print (June 2022), 8. https://doi.org/10.1097/CEH.0000000000000444

[38] H. Quan, B. Li, C. M. Couris, K. Fushimi, P. Graham, P. Hider, J.-M. Januel, and V. Sundararajan. 2011. Updating and Validating the Charlson Comorbidity Index and Score for Risk Adjustment in Hospital Discharge Abstracts Using Data From 6 Countries. *American Journal of Epidemiology* 173, 6 (March 2011), 676–682. https://doi.org/10.1093/aje/kwq433

[39] Ayanna Seals, Giuseppina Pilloni, Jin Kim, Raul Sanchez, John-Ross Rizzo, Leigh Charvet, Oded Nov, and Graham Dove. 2022. 'Are They Doing Better In The Clinic Or At Home?': Understanding Clinicians' Needs When Visualizing Wearable Sensor Data Used In Remote Gait Assessments For People With Multiple Sclerosis. In *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–16. https://doi.org/10.1145/3491102.3501989

[40] Tim Shaw, Anna Janssen, Roslyn Crampton, Fenton O'Leary, Philip Hoyle, Aaron Jones, Amith Shetty, Naren Gunja, Angus G Ritchie, Heiko Spallek, Annette Solman, Judy Kay, Meredith AB Makeham, and Paul Harnett. 2019. Attitudes of health professionals to using routinely collected clinical data for performance feedback and personalised professional development. *Medical Journal of Australia* 210, S6 (April 2019), 5. https://doi.org/10.5694/mja2.50022

[41] John Szakiel. 2010. Hospital Casemix Protocol — Medibank Private Perspective. *Health Information Management Journal* 39, 2 (June 2010), 47–48. https://doi.org/10.1177/183335831003900209

[42] Maryam Tohidi, William Buxton, Ronald Baecker, and Abigail Sellen. 2006. Getting the Right Design and the Design Right: Testing Many Is Better Than One. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*. ACM, New York, NY, United States, 10.

[43] Konstantina Vasileiou, Julie Barnett, Susan Thorpe, and Terry Young. 2018. Characterising and justifying sample size sufficiency in interview-based studies: systematic analysis of qualitative health research over a 15-year period. *BMC Medical Research Methodology* 18, 1 (Dec. 2018), 148. https://doi.org/10.1186/s12874-018-0594-7

[44] Vivian L West, David Borland, and W Ed Hammond. 2015. Innovative information visualization of electronic health record data: a systematic review. *Journal of the American Medical Informatics Association* 22, 2 (March 2015), 330–339. https://doi.org/10.1136/amiajnl-2014-002955

[45] Huaqiong Zhou, Phillip R Della, Pamela Roberts, Louise Goh, and Satvinder S Dhaliwal. 2016. Utility of models to predict 28-day or 30-day unplanned hospital readmissions: an updated systematic review. *BMJ Open* 6, 6 (June 2016), e011060. https://doi.org/10.1136/bmjopen-2016-011060

[46] Torre Zuk and Sheelagh Carpendale. 2007. Visualization of Uncertainty and Reasoning. In *Smart Graphics*, Andreas Butz, Brian Fisher, Antonio Krüger, Patrick Olivier, and Shigeru Owada (Eds.). Vol. 4569. Springer Berlin Heidelberg, Berlin, Heidelberg, 164–177. https://doi.org/10.1007/978-3-540-73214-3_15 ISSN: 0302-9743, 1611-3349 Series Title: Lecture Notes in Computer Science.

A APPENDIX

A.1 Prototype data visualisation designs

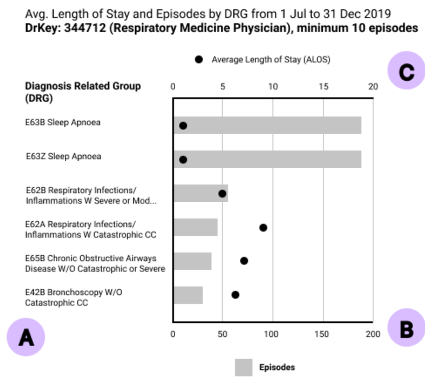


Fig. 1. **Bar chart.** Shows **(A)** the DRG; a classification system which provides clinically meaningful ways to relate a number and type of patients treated in a hospital to the resources required [33]; **(B)** the Number of Patient Episodes; a period of admitted patient care between admission to separation e.g. discharge, transfer, or death; and **(C)** the Average LOS; the mean number of days patients stay overnight in the hospital.[1]

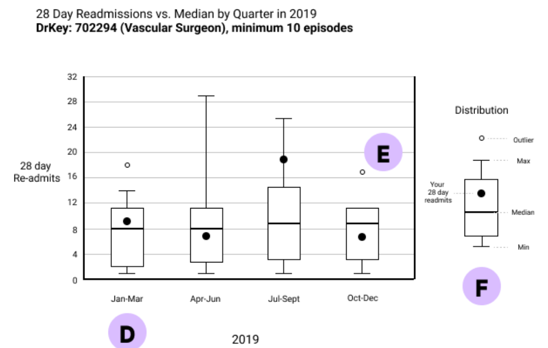


Fig. 2. **Box plot.** Shows **(D)** the distribution of 28-day readmissions for all Vascular Surgeons in the January-March quarter; **(E)** the outlier surgeons where their 28-day readmissions fall outside of the whiskers (minimum or maximum); and **(F)** the legend to show how an individual surgeon compares with the cohort median, minimum, maximum, and outliers.[1]

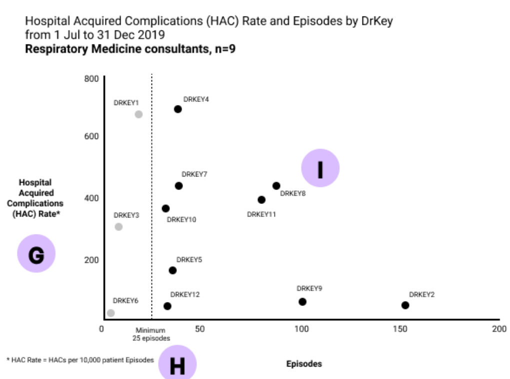
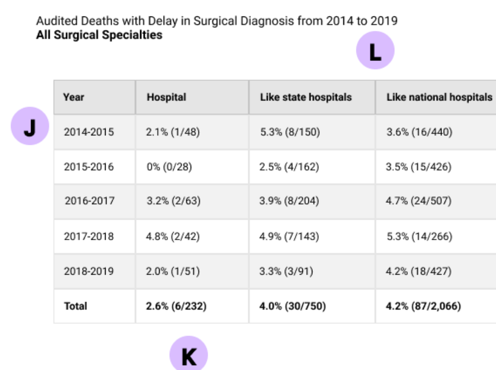


Fig. 3. **Scatter plot.** Shows **(G)** the HAC Rate, Hospital Acquired Complications are patient complications for which clinical risk mitigation strategies may reduce (but not necessarily eliminate) the risk of that complication occurring [4] e.g. falls, infections, renal failure, and cardiac complications; **(H)** the Minimum Number of Patient Episodes cut-off; and **(I)** the unique identifier for each clinician. [1]



Year	Hospital	Like state hospitals	Like national hospitals
2014-2015	2.1% (1/48)	5.3% (8/150)	3.6% (16/440)
2015-2016	0% (0/28)	2.5% (4/162)	3.5% (15/426)
2016-2017	3.2% (2/63)	3.9% (8/204)	4.7% (24/507)
2017-2018	4.8% (2/42)	4.9% (7/143)	5.3% (14/266)
2018-2019	2.0% (1/51)	3.3% (3/91)	4.2% (18/427)
Total	**2.6% (6/232)**	**4.0% (30/750)**	**4.2% (87/2,066)**

Fig. 4. **Table.** Shows **(J)** the time period, a calendar year; **(K)** the percentage and count of Deaths with Delayed Surgical Diagnosis; and **(L)** benchmarked data for comparable state and national hospitals.[1]

[1]Data points in the figures have been adjusted prior to publication to preserve the confidentiality of the hospital sites from which the datasets were sourced.