Feedback of aggregate patient-reported outcomes (PROs) data to clinicians and hospital end users: findings from an Australian codesign workshop process

Olivia Francis Ryan,1,2 Shaun L Hancock,1,2 Violet Marion,1 Paulette Kelly,3 Monique F Kilkenny,1,3 Benjamin Clissold,4,5 Penina Gunzburg,6 Shae Cooke,7 Lauren Guy,8 Lauren Sanders,9,10 Sibilah Breen,1 Dominique A Cadilhac1,3

ABSTRACT

Objectives Patient-reported outcomes (PROs) are increasingly used to measure the patient’s perspective of their outcomes following healthcare interventions. The aim of this study was to determine the preferred formats for reporting service-level PRO data to clinicians, researchers and managers to support greater utility of these data to improve healthcare and patient outcomes.

Setting Healthcare professionals receiving PRO data feedback at the health service level.

Participants An interdisciplinary Project Working Group comprised of clinicians participated in three workshops to codesign reporting templates of summarised PRO data (modified Rankin Scale, EuroQol Five Dimension Descriptive System, EuroQol Visual Analogue Scale and Hospital Anxiety and Depression Scale) using a modified Delphi process. An electronic survey was then distributed to shortlist the preferred templates among a broad sample of clinical end users. A final workshop was undertaken with the Project Working Group to review results and reach consensus on the final templates.

Primary and secondary outcome measures The recommendation of preferred PRO summary data feedback templates and guiding principles for reporting aggregate PRO data to clinicians was the primary outcome. A secondary outcome was the identification of perceived barriers and enablers to the use of PRO data in hospitals. For each outcome measure, quantitative and qualitative data were summarised.

Results 31 Working Group members (19 stroke, 2 psychology, 1 pharmacy, 9 researchers) participated in the workshops, where 25/55 templates were shortlisted for wider assessment. The survey was completed by 114 end users. Strongest preferences were identified for bar charts (37/82 votes, 45%) and stacked bar charts (37/91 votes, 41%). At the final workshop, recommendations to enhance communication of PROs data for comparing health service performance were made including tailoring feedback to professional roles and use of case-mix adjustment to ensure fair comparisons.

Conclusions Our research provides guidance on PROs reporting for optimising data interpretation and comparing hospital performance.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ We used iterative codesign methods with interdisciplinary clinical or policy stakeholders who were considered the end users of the evidence we were to generate from this study. A modified Delphi process was used to reach consensus on the PRO summary data feedback templates and to establish guidance for the preferred reporting formats.

⇒ Strengths of our study include the engagement of clinical, academic, patient and government representatives who were members of the Working Group for the Delphi process and involvement of a wider group of stakeholders who completed an electronic survey.

⇒ The use of different commonly used PRO measures was a strength, as we trialled both condition-specific and generic PRO measures in an effort to increase the generalisability of our results across different clinical populations.

⇒ A limitation was our inability to include examples of the presentation of baseline and follow-up PRO data due to the sample data being derived from the stroke registry where PROs are only collected at a single time point after the acute event. We acknowledge that these data are important for a range of conditions such as chronic kidney disease or surgical interventions where progression may be important to monitor.

⇒ Due to the use of the condition of stroke as a case study our findings may lack generalisability for other clinical conditions. However, the principles are largely transferable. The next steps are to pilot and refine the recommended templates.

INTRODUCTION

Patient-reported outcomes (PROs) are data collected directly from the patient about their own health, without interpretation by a clinician or any other person.1 The use of PROs in comparative-effectiveness research and clinical trials is well established.2 There is increased recognition that PROs have the
potential to enhance clinical practice through directing the need for improvements in the safety and quality of healthcare.3–5

PRO data can be used at the individual-patient level as a means to inform clinicians on aspects of their health that are important to them6 or at the health service level to enable clinicians to compare hospital performance or against achievable benchmarks to identify quality improvement opportunities.7 8 The interpretation of aggregate, health service-level PROs data by clinicians and other hospital end users is the focus of our research.

Known challenges related to the interpretation and use of PRO data include scoring and scaling differences between measures as well as inconsistencies in the methods used to report PRO data back to clinicians.9 To address these challenges, our preliminary research involved a scoping review to summarise the existing evidence related to preferred formats for PRO data feedback to clinicians.10 While a single preferred format or approach to feedback PRO data to healthcare professionals was not identified, we could summarise some general guidance on how to design feedback formats to enhance the interpretation of summarised PRO data.

The aim of the current study was to build on our preliminary research and establish guidance for preferred formats for hospital comparisons of PRO data to support greater utility of these data to improve healthcare and patient outcomes. In terms of quality of care, PROs can have relevance to clinicians if these data are considered in addition to information about performance on standards of clinical care (eg, expected interventions to be provided) and other information such as mortality and readmissions. For example, multicomponent audit and feedback approaches whereby PROs data are used to complement clinical indicator data can have an additive effect and contribute to better care through improving adherence to clinical guidelines.11 Other authors12 have acknowledged that there is currently a lack of primary studies and randomised controlled trials related to the effectiveness of aggregate PRO data for benchmarking and quality improvement purposes and the associated impact on patient outcomes.13 Despite this, PRO feedback is an emerging field and our research constitutes an important development in this area. We also sought to gain preliminary insights into the perceived barriers and enablers to the use of PRO data in hospitals.

METHODS
Study design
This was a codesigned project, whereby the methods included several interdisciplinary end users working together to produce the outcomes.14 We defined end users as hospital clinicians from a range of disciplines, academics who analyse and report PRO data and government representatives who use PRO data to inform quality improvement and policy decisions. We used a modified Delphi process15 to reach consensus on PRO summary data templates and establish guidance for preferred reporting formats. The Delphi method is a group communication process, which aims to achieve convergence of opinions on a specific real-world issue.15 The current study was designed in three stages: (1) a series of three workshops whereby members of a Project Working Group (herein referred to as Working Group) iteratively refined several PRO data feedback templates, (2) an anonymous electronic survey disseminated widely across Australia via clinical and academic networks to enable the short listing of the codesigned PRO summary data templates, (3) a final workshop with the Working Group members to achieve consensus on the recommended set of PRO summary data reporting format characteristics (figure 1).

Figure 1  Diagrammatic representation of project stages. EOI: expression of interest; WS: workshop
Setting
This study was commissioned by the Victorian Agency for Health Information (VAHI) as part of a broader strategy to facilitate the comprehensive collection of PRO data in Victoria, Australia. To inform the state-wide strategy, VAHI commissioned consultancy work by Paxton Partners in 2018, whereby the authors identified that regular feedback of PRO data had the potential to add substantial value to improve clinical engagement with these data (Paxton Partners, PRO measures (PROMs): Literature Scan, personal communication, 2019). In response, VAHI appointed a project team comprising representatives from The Florey, Safer Care Victoria and Monash University to build on this work (see online supplemental file 1).

Currently, published evidence related to the systematic use of aggregated PRO data for quality improvement initiatives and policy decisions among clinicians and other stakeholders, such as government representatives, is sparse. The increasing ability to collect PROs and the proliferation of Clinical Quality Registries internationally,16 has increased the need to establish rigorous evidence regarding the best feedback methods and preferred graphical formats for reporting PROs data to clinicians and other stakeholders. This current knowledge gap exists even though there is a strong policy rationale to support high functioning, mature Clinical Quality Registries with the ability to feedback data (both clinical and PROs)17 and despite the fact that promoting the collection of PRO data is published as a priority for Australia in the 2020–2025 National Health Reform Agreement.18

The current study was initiated to assess the perceived barriers and opportunities for uptake of PRO data among clinicians and to establish best practice feedback approaches to be adopted by Clinical Quality Registries. Since 2012, PROMs (generic and stroke-specific) have been collected by The Florey as part of its role in data management processes of the Australian Stroke Clinical Registry (AuSCR).19 Therefore, the condition of stroke was chosen as a case study.

Participants
The Working Group was established following the circulation of an expression of interest by the Victorian Stroke Clinical Network. Our objective was to achieve broad, interdisciplinary representation of experts and those with direct interests in the transparent feedback of PRO data. Individuals who were based in metropolitan or regional areas in Victoria and working in stroke care from acute, subacute and community settings were eligible to self-nominate for membership. We also included a consumer representative who was a survivor of stroke and had experience working with the Victorian government on various committees (see online supplemental file 1).

Stage 1: codesign workshops
In stage 1, we conducted three workshops between August and October 2019, with the option for participants to attend face-to-face or via the videoconferencing platform Zoom.20 The meetings were chaired by authors DAC/SB and the content discussion, for example, data templates were facilitated by an experienced moderator (VM). The first workshop oriented the Working Group members to the aims of the current study and was used to provide an outline of the general reporting principles from the existing literature.20 The identification of priorities and opinions was facilitated through the use of open-ended questions (eg, ‘Who has used PROs data reported for their health service?’, ‘Do you think PROs data add value, why/why not?’). At the conclusion of the first workshop, it was agreed that the Florey project team (authors OFR, SLH, VM, SB and DAC) would develop a set of templates based on three PROMs commonly used in clinical practice. These included: the modified Rankin Scale (mRS),21 EuroQol Five Dimensions Three Levels (EQ-5D-3L)22, and the Hospital Anxiety and Depression Scale (HADS)23 (see online supplemental table 1). The PROMs were pragmatically chosen based on the experience of the project team with input from VAHI representatives.24 All templates used fictitious hospital, state and/or national-level data. For the templates designed to facilitate benchmarking of PRO data, peer hospitals were defined as those facilities with a similar bed size/number of admissions per year and population served. For these types of templates, the data from all eligible peer hospitals were included as a comparison. The number of peer hospitals is specified in the relevant graph axis label, figure legend or instructional footnote.

The initial set of templates developed after the first workshop were circulated to the Working Group prior to the second workshop, via an electronic survey developed using SurveyMonkey software.25 The Working Group members then ranked the templates according to preference. During the second workshop, members discussed the results of the survey and critiqued the templates to identify the prevailing preferences using Zoom live polling.26 Refinements to the templates were made, then circulated via an electronic survey prior to the third workshop. The third workshop mirrored that of the second, whereby the survey results were discussed and members could revise their preferences. Agreement was reached for the final set of templates to be shortlisted in stage 2 (see online supplemental files 2–9). All workshops were audio recorded. At the conclusion of each workshop, formal minutes and the aggregated survey results were circulated to all members.

Stage 2: anonymous electronic survey
An anonymous electronic survey, from a broad sample of potential PRO data end users who had not been part of stage 1, was used to evaluate the feedback templates finalised in stage 1. The survey was designed using SurveyMonkey software.25 Survey respondents were recruited using a convenience sampling approach via invitations sent from several relevant clinical and academic networks. The survey link included instructions and outlined the
voluntary participation requirements. Eligibility was self-determined by individuals who identified as a healthcare professional or academic/clinical researcher interested in the use of PRO data. Where feasible, snowball sampling was also encouraged, whereby respondents could share the link with other colleagues. The survey remained open for four weeks. Reminders to the main dissemination groups were sent one week prior to the closure of the survey.

The survey included questions on the characteristics of respondents, as well as feedback on each of the PRO templates for each specific measure, including mRS, EQ-5D-3L and the HADS. A set of templates that presented PRO data over time were also included. Using previous PRO-related stakeholder survey methods reported by Brundage et al., a series of scale response questions (0=least, 10=most) were asked to ascertain: (a) the respondents’ rating of the perceived utility: ‘How useful do you find this graph?’; (b) the respondents’ rating of their ability to understand each sample format: ‘How easy it is for you to understand this graph?’. The respondents then selected the preferred template out of all the options for each category and could provide a free-text comment. Multiple-choice and open-ended questions were included to elicit perceived barriers and enablers related to the use of PRO data. There were three survey questions specific to barriers and enablers, for example, ‘What are the main barriers to the use of PROMs in your clinical practice?’.

### Stage 3: final consensus workshop
The Working Group was reconvened for a fourth (final) workshop to review the results of the anonymous electronic survey. The workshop discussion was guided by the same moderator (VM) using open-ended questions to achieve a final consensus on the recommended PRO feedback templates (eg, ‘Could this graph be modified?’; ‘Do you agree this format is the preferred template for this PROM, why/why not?’).

### Statistical analysis
The survey data collected as part of stage 1 (preworkshop and live polls) and stage 2 (electronic survey) were extracted and summarised using Microsoft Excel (V.2016). The quantitative data were summarised using descriptive analyses (totals and proportions). The stage 2 anonymous survey scale data were summarised as medians and IQRs. The stage 2 survey data were analysed overall, as well as according to clinical versus non-clinical respondents. Clinical respondents were defined as those with a medical, nursing or allied health background. Non-clinical respondents were defined as health system administrators, policymakers or researchers. The free-text qualitative data collected as part of the stage 2 survey were analysed using thematic analysis. An inductive approach was used, whereby one member of the project team identified broad theme categories related to each question (VM). A second reviewer (OFR) cross checked the qualitative responses to verify the themes. Both the qualitative and quantitative data were used with the objective of determining prevailing preferences and preferred formatting attributes.

### Patient and public involvement
The main consumer group for this research were clinicians for which the data summary reports are focused. A patient consumer representative was included as part of the Working Group for this project. No other public representatives or patients were involved in the overall study. Clinician representatives were asked to advise on the interpretation and write up of the results, and several members of our working group were included as co-authors on this article.

### RESULTS
The Working Group comprised 33 members (19 clinicians, 1 consumer representative, 13 project team members) (table 1). The majority of clinical representatives were women (74%) and from an allied health (42%) or nursing (26%) profession. Three members were from other professional backgrounds, including psychology and pharmacy. Working Group members from a clinical background had a median of 12.5 years in clinical practice. Stroke management was the field of expertise for 84% (n=16) of the clinical representatives.

### Stage 1: codesign workshops
Overall, 31 Working Group members actively participated in one or more workshops (see online supplemental file 1). The number and type of templates presented during each workshop are summarised in online supplemental table II. The templates were refined or retired based on the preferences identified during each workshop.

| Table 1 Demographic characteristics of the project working group (N=33)* |
|-----------------------------|-----------------------------|-----------------------------|
| Characteristic              | Clinicians N=19 n (%)       | Project team N=13 n (%)     |
| Male                        | 5 (26)                      | 2 (15)                      |
| Profession                  |                             |                            |
| Doctor                      | 4 (21)                      | –                           |
| Nurse                       | 5 (26)                      | –                           |
| Allied health               | 8 (42)                      | –                           |
| Other†                      | 2 (11)                      | 13 (100)                    |
| Clinical area of expertise  |                             |                            |
| Stroke                      | 16 (84)                     | 13 (100)                    |
| Other                       | 3 (16)                      | –                           |
| Median years in clinical practice (Q1, Q3) | 12.5 (8, 15) | – |

*Includes one male consumer representative.
†Other: refers to professional backgrounds including psychology and pharmacy.
The formatting preferences included: simple layouts or symbols to reduce complexity, use of definitions and instructions (where appropriate), use of normative population data, CIs and denoting sample sizes (see online supplemental table III). Several templates were determined to be unsuitable for service-level PRO reporting, including spider plots, heat maps and pictographs. The Working Group agreed these formats were more complex and would limit the ability to readily interpret the results, while pictographs were considered too simplistic with insufficient detail for this audience. Following the third workshop, a separate category containing templates used for reporting PROs data over time was supported for wider testing, resulting in six additional templates.

Stage 2: anonymous electronic survey

There were 114 respondents for this survey (table 2). The majority of respondents were women (76%) and aged between 30 and 49 years (56%). Respondents were predominantly clinical representatives (91%), including allied health (38%) or nursing (32%). Overall, 87 respondents (71%) reported stroke as their area of clinical expertise. The respondents had a median of 20 years in clinical practice, and 23% reported using PROs at the time of the survey. The following sections detail the results for each PRO category assessed as part of the survey. The denominators vary since not all respondents completed every question used to evaluate the templates. The overall response rate for the survey was 61%.

Preferences for graphical formats of summary data

When interpreting mRS data, overall, 41% (n=37/91) of survey respondents preferred format A (see online supplemental files 2 & 3) which reflected a two-bar horizontal stacked bar chart with comparative data (table 3). The prevailing template had the greatest rating for ability to understand the data being presented (median=9), despite only a two-point median difference between the five templates presented (see online supplemental table IV). The responses of non-clinical respondents were similar to clinical respondents, with a preference for an additional table with mRS data values paired with the stacked bar chart.

For EQ-VAS data, 45% of respondents (37/82) preferred format A (see online supplemental file 4) comprised of a bar chart with the inclusion of CIs and stratification by patient sex and age group. Compared with the more complex formats (eg, format C: caterpillar plot), the bar chart resulted in the greatest rating (median=7) for both perceived ease of understanding and usefulness (see online supplemental table IV). The median ratings across the three EQ-VAS templates were similar, with a maximum median difference of two points only. The responses of non-clinical respondents differed to clinical respondents, with a preference for a dot or caterpillar plot.

When interpreting EQ-5D-3L data, format F (see online supplemental files 5 & 6), which comprised a stacked bar chart presenting data for all five dimensions (with categorisation according to the proportion of patients reporting ‘no problems’ vs those reporting ‘problems’), was identified as the preferred format (22/78, 28%). Despite the greatest proportion of respondents preferring format F, there was a diverse range of median ratings across all seven templates. Several anomalies were identified whereby four other templates (eg, formats A and B with vertical stacked bar chart and traffic light colour
the ease of understanding, which was greater than that of Ryan OF, et al. BMJ Open 2022;12:e055999. doi:10.1136/bmjopen-2021-055999

...to depict a clinical cut-off point, which indicated possible anxiety or depression (see online supplemental file 7).

A total of 21/69 (30%) respondents preferred format E (see online supplemental files 8 & 9), a line graph template for presenting data over time. The line graph that was short listed as the preference displayed hospital data and comparative state-level hospital data, with additional metrics to stratify results according to patient characteristics (e.g., stroke type) as well as a dotted line to indicate normative population estimates. Although this template prevailed, there were two other variations of a line graph (formats A and B) that had identical median ratings (understanding=8, usefulness=7), along with two bar graphs (formats C and D). The consistent median ratings across all formats did not align with the differences observed between the overall rankings of preference. There was no difference for the overall preference between clinical and non-clinical respondents.

Several themes emerged from the qualitative data related to formatting preferences, including: the use of colours, particularly intuitive colours such as green=better/good outcome and red=worse/poorer outcome (see online supplemental table III). Though, the use of dark/shaded colours was cautioned to ensure templates would be interpretable in grayscale (e.g., to facilitate printed versions for hospital end users). In addition, the inclusion of clear...
labelling (namely graph axes and titles), sample sizes and proportions were also identified as preferred features. Though, the inclusion of too much explanatory text was cautioned. The respondents considered stratification of PRO summary data by age, sex and other variables to be useful, particularly to assist with targeted quality improvement strategies.

The inclusion of benchmarked peer-level, state-level and/or national-level data was also supported to facilitate hospital performance monitoring. The respondent’s qualitative responses identified that benchmarks provided information about targets while peer data allowed for comparison against similar hospitals and within jurisdictions. The respondents also preferred the term ‘peer hospitals’ to be clearly defined. The need to provide case-mix adjusted data was identified by several respondents across various templates, with the importance of this statistical adjustment highlighted for the presentation of longitudinal PRO data (eg, otherwise variation could be due to the difference in patient characteristics each year).

**PRO data reporting interval preferences and perceived barriers and enablers**

The majority of respondents (60/69, 87%) preferred a 3 to 6-month interval for PRO data reporting. The main perceived barrier to using PRO data in clinical practice from the summarised qualitative data, included resource and time constraints (35/68, 51%) (including staff time to read and interpret the information). Other perceived barriers included a lack of tools available to facilitate meaningful reporting of PROs data, a lack of understanding about how to use the information and lack of organisational support to use the data or a perceived need for culture change. Two of the most common themes identified as enabling factors included: an interest to use PROs as a mechanism to deliver patient-centred care, and the development of enhanced feedback methods to facilitate greater PRO data uptake and future use. Other reported enablers included the need for adequate resources such as time and funding for clinicians to interpret and use the data and enhanced education about how best to interpret and use the data to improve patient care.

**Stage 3: Final consensus workshop**

The Working Group agreed with the outcomes of the stage 2 survey results and the shortlisted templates. The median ratings for each template, especially for formats where anomalies, were identified in contrast to the overall preference, were used as a supplement for the Working Group to consider. Figure 2 illustrates two templates recommended for presenting: data at a single time point versus data over time, along with a summary of preferred formatting features. An example of a template that was identified to not be appropriate for routine feedback purposes is presented in figure 3. The Working Group agreed that the professional roles of the recipients of these data would require consideration, and that certain recipients (eg, hospital executives) may require further detail such as a pairing a graphical display with a table of data values. The Working Group discussed several data quality aspects of PROs (eg, the need for case-mix adjustment) which were supported for incorporation into the final project recommendations (see online supplemental table V).

**DISCUSSION**

PROs data have the potential to improve healthcare and patient outcomes, but in order for this to occur, it is essential that PROs data are presented in a manner that is both useful and understandable. The presentation of aggregated PROs data is often done in the absence of best practice guidance. Guided by evidence from our previous literature review, this study aimed to use a codesign process to create a set of PRO summary data reporting guidelines.
templates for health service-level reporting. The recommended templates and preferred formatting attributes identified from our study complement existing mechanisms proposed by Snyder et al. and contribute much needed evidence for methods to use when communicating service-level PRO data to clinicians.

Despite some mixed opinions among the clinical representatives related to each template, a clear preference emerged to support the provision of these data. Our findings indicate that the simpler formats reduced the cognitive burden for clinicians, and there was a preference for use of explanatory attributes to aid interpretation of the data. For data presented at a single time point, a bar or stacked bar chart with up to four bars prevailed. A caveat to this recommendation was for PROMs with multiple dimensions, such as the EQ-5D, where clinicians preferred to review local and comparison hospital data within the one graph (up to 10 bars). A line graph with a maximum of four lines and minimal additional constructs was identified as the preference for data presented over time. A construct can include statistical or formatting features such as: stratification according to patient demographic/clinical characteristics or providing a reference to normative population data or clinical cut-offs. Our main findings are consistent with the preliminary principles for graphical display of PRO data proposed by Bantug et al. and builds on the ‘less is more’ guidance recommended for effective communication of performance monitoring or clinical data in general.

The diversity of opinions among clinicians in the Working Group and broader survey sample in our study is not dissimilar to previous research. Brundage et al. evaluated the preferences of graphically displayed PRO results among 233 clinicians and found differing preferences for bar and pie charts. In the current study, we also found that clinicians had differing preferences related to the use of pie charts compared with bar charts as well as intuitive colour coding schemes (eg, some clinicians indicated that traffic light colours added visual clarity, whereas others disliked them). Furthermore, the discrepancies identified from the median ratings in this current study, in which several templates received a rating of >8 despite not being identified as the prevailing format, highlight the challenge of recommending a single format style for all target audiences. Despite some mixed views, and mostly representation from professionals working in the clinical area of stroke, the recommendations from this current study form an empirical basis for future work in this field. We recommend that there is a need to involve hospital end users, including representatives from quality improvement departments, if aggregate PRO data are being prepared by academics, and that these data should be presented in more than one way to address the needs of specific audience subgroups. The barriers and enablers identified in this current study, including resource and time constraints and education for clinicians, align with the findings of our previously conducted scoping review into PRO feedback.

Figure 3 Example illustrating a non-preferred template for reporting aggregate, service-level patient reported outcomes data to clinicians.
We also found a preference for the majority of the shortlisted templates to include comparative hospital data. In our examples, we followed the approach used in the AuSCR to provide peer-hospital, state-level and national-level benchmarks. Case-mix adjustment for PROs data is an emerging field and the application of case-mix methods is a prerequisite for quality reporting and benchmarking purposes. Further empirical research is needed to identify patient-related (eg, age, sex, life events, new healthcare episodes) and hospital-related (workload, volume of patients, hospital type, etc) factors requiring adjustment for reporting benchmarked PRO data. It was intriguing that only a single survey respondant considered the importance of case-mix adjustment, and this might suggest that education for clinicians about the importance of case-mix adjustment when reviewing comparisons of PROs at the service level be undertaken. This current study contributes to the discourse in confirming the need for ongoing methodological research to determine the most appropriate analytic methods and the variables needed to enable reliable case-mix adjustment. In the interim, it is recommended that when providing descriptive PROs comparisons without the ability to adjust for differences in patient characteristics, a concise and explicit definition of what constitutes a peer hospital must be included in the template.

Strengths and limitations

The strengths of our study include the engagement of an interdisciplinary sample of clinical, academic and government representatives who were members of the Working Group, and other potential end users who completed the electronic survey. This is one of only a few studies in the field, outside of a cancer setting, to iteratively develop and evaluate the preferences of clinicians when considering the use of PROs for comparing hospital performance. Applying a modified Delphi process throughout each stage was also a strength as well as the use of multiple engagement tools to optimise participation of Working Group members (eg, live polling). An additional strength was the use of different commonly used PROs measures as part of the templates. While previous work has focused on the development of templates for PRO measures for a single condition, we trialled both condition-specific and generic PRO measures in an effort to increase the generalisability of our results across different clinical populations.

The limitations of our study include that our templates were based on experience reporting PROs as part of the AuSCR and informed by the literature review. Some subjective interpretation of how to display data and how many templates to produce for assessment was unavoidable. Although we sought broad representation for the survey by advertising widely, we acknowledge that the majority of the Working Group were experienced in the field of stroke and were located within a single jurisdiction. A further limitation was our inability to include examples of the presentation of baseline and follow-up PRO data due to the sample data being derived from the stroke registry where PROs are only collected at a single time point after the acute event. We acknowledge these data are important for a range of conditions such as chronic kidney disease or surgical interventions, where progression may be important to monitor. We acknowledge the presence of responder bias in our survey data and that the ability to detect preferences among different professional audience groups (eg, medical/nursing staff vs government representatives/researchers) was limited by the composition of mostly female respondents who self-selected to participate in the survey. Despite this, our findings align with those from preliminary work conducted in other patient populations, which suggests that there are general principles for the presentation of aggregate PRO data that are applicable across clinical specialities. We acknowledge that the next steps are to pilot and refine the recommended templates and encourage others to consider evaluating these as well.

CONCLUSIONS

We have illustrated the iterative process and outcomes of a codesigned approach to establishing summary data templates for reporting aggregate service-level PROs data to clinicians. Simple graphical templates, with accompanying instructions and formatting attributes to aid data interpretation, were identified as the preferred format characteristics. This work provides important evidence for Clinical Quality Registries and other organisations that routinely feedback aggregate PRO data, to ensure that the potential of these data to support quality improvement efforts is fully realised.

Author affiliations

1Public Health: Stroke Division, Florey Institute of Neuroscience and Mental Health—Austin Campus, Heidelberg, Victoria, Australia
2Victorian Agency for Health Information, Victoria Department of Health and Human Services, Melbourne, Victoria, Australia
3Stroke and Ageing Research, Department of Medicine, School of Clinical Sciences at Monash Health, Monash University, Clayton, Victoria, Australia
4Neurosciences Department, Monash Health, Clayton, Victoria, Australia
5Neurosciences Department, Barwon Health, Geelong, Victoria, Australia
6Physiotherapy Department, Alfred Health, Melbourne, Victoria, Australia
7Department of Physiotherapy, Eastern Health, Box Hill, Victoria, Australia
8Community Based Rehabilitation, Sunshine Hospital, Saint Albans, Victoria, Australia
9Department of Neurosciences, St Vincent’s Hospital Melbourne Pty Ltd, Fitzroy, Victoria, Australia
10Department of Medical Education, The University of Melbourne - Parkville Campus, Melbourne, Victoria, Australia

Twitter Monique F Kilkenney @KilkenneyMonique and Dominique A Cadilhac @DominiqueCad

Acknowledgements We acknowledge the members of Patient Reported Outcome Measures Working Group (see Supplementary 1). We also thank the hospital clinicians, executive and administrative staff, policy makers, consumer representative, researchers and clinical registry special interest groups who contributed to the anonymous electronic survey (see Supplementary 1). We acknowledge the contribution of hospital staff from sites for collecting data based on patients registered in the AuSCR.
REFERENCES


Feedback of aggregate, health-service level patient-reported outcomes data

SUPPLEMENTAL MATERIAL

Feedback of aggregate Patient-Reported Outcomes (PROs) data to clinicians and hospital end-users: Findings from an Australian co-design workshop process

Olivia F Ryan¹ (0000-0003-4977-6742),
Shaun Hancock¹ (0000-0002-2015-2752),
Violet Marion¹ (0000-0001-6643-7035),
Paulette Kelly²,
Monique F Kilkenny¹,³ (0000-0002-3375-287X)
Benjamin Clissold⁴,⁵
Penina Gunzburg⁶ (0000-0002-6108-0302)
Shae Cooke³ (0000-0001-5531-0135)
Lauren Guy⁸ (0000-0003-1143-5117)
Lauren Sanders⁹,¹⁰
Sibilah Breen¹ (0000-0001-9896-004X)
Dominique A Cadilhac¹,³ (0000-0001-8162-682X)

Affiliations:

1. Stroke Theme, the Florey Institute of Neuroscience and Mental Health, Heidelberg, VIC, Australia
2. Victorian Agency for Health Information, Department of Health and Human Services, Victorian government, Australia
3. Stroke and Ageing Research, Department of Medicine, School of Clinical Sciences at Monash Health, Monash University, Clayton, VIC, Australia
4. Neurosciences Department, Monash Health, VIC Australia
5. Neurosciences Department, Barwon Health, VIC Australia
6. Alfred Health Physiotherapy, VIC, Australia
7. Department of Physiotherapy, Eastern Health, VIC, Australia
8. Community Based Rehabilitation, Sunshine Hospital, Western Health, St Albans, VIC, Australia
9. Department of Neurosciences, St Vincent’s Hospital, Melbourne, Fitzroy, VIC, Australia
10. Department of Medical Education, University of Melbourne, Parkville, VIC, Australia

Supplemental Methods

Table I. Overview of the Patient-Reported Outcome Measures used for developing feedback formats

Supplemental Results

Table II. Number and type of feedback templates tested during each project stage

Table III. Summary of outcomes according to each project stage

1
Feedback of aggregate, health-service level patient-reported outcomes data

Table IV: Median ratings for the templates tested in the online survey (stage two)

Table V: Recommendations to guide feedback of aggregate, summary Patient Reported Outcomes data to hospital end-users

Supplemental Acknowledgments: Patient Reported Outcome Measures Working Group and other contributors

Supplementary Files

File 2: modified Rankin Scale templates (Format A to D)
File 3: modified Rankin Scale templates (continued) (Format E)
File 4: EQ-VAS templates (Format A to C)
File 5: EQ-5D-3L descriptive system data templates (Format A to D)
File 6: EQ-5D-3L descriptive system data templates continued (Format E to G)
File 7: HADS templates (Format A to D)
File 8: Longitudinal data templates (Format A to D)
File 9: Longitudinal data templates (Format E to F)
**Feedback of aggregate, health-service level patient-reported outcomes data**

### Supplemental Methods

Table I. Overview of the Patient-Reported Outcome Measures used for developing feedback formats

<table>
<thead>
<tr>
<th>Patient-Reported Outcome Measure</th>
<th>Reference</th>
<th>Brief description of survey instrument</th>
<th>Outcomes measured</th>
<th>Number of survey items</th>
<th>Generic or condition-specific measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modified Rankin Scale (mRS)</td>
<td>Rankin J. Cerebral vascular accidents in patients over the age of 60. Scott Med J 1957;2:200-15</td>
<td>A scale to measure the degree of disability or dependence in the daily activities of people who have experienced a stroke or other causes of neurological disability</td>
<td>The patient's functional status recorded on an ordinal scale with 6 categories:</td>
<td>6</td>
<td>Condition-specific</td>
</tr>
</tbody>
</table>
## Feedback of aggregate, health-service level patient-reported outcomes data

<table>
<thead>
<tr>
<th>Patient-Reported Outcome Measure</th>
<th>Reference</th>
<th>Brief description of survey instrument</th>
<th>Outcomes measured</th>
<th>Number of survey items</th>
<th>Generic or condition-specific measure</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hospital Anxiety and Depression Scale (HADS)</strong></td>
<td>Snaith RP. The Hospital Anxiety And Depression Scale. <em>Health and Quality of Life Outcomes.</em> 2003;1:29</td>
<td>A screening tool for anxiety and depressive symptoms that is not dependent on physical symptoms. Has two 7-item subscales for anxiety and depression. Each item is recorded on an ordinal scale with 4 categories from zero (least distress) to three (greatest distress). The scale scores are calculated by summing the responses to the items up to a maximum score of 21 points per sub-scale and out of 42 points in total.</td>
<td>Anxiety Depression</td>
<td>14</td>
<td>Generic</td>
</tr>
</tbody>
</table>

*The modified Rankin Scale category of ‘6’ can be added to signify death.*
Feedback of aggregate, health-service level patient-reported outcomes data

Supplemental Results

Table II. Total number of feedback templates tested during each project stage.

<table>
<thead>
<tr>
<th>Template category</th>
<th>Stage 1: Co-design workshop process*</th>
<th>Stage 2: Online survey</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Workshop 2</td>
<td>Workshop 3</td>
</tr>
<tr>
<td></td>
<td>Formats presented</td>
<td>Formats retired</td>
</tr>
<tr>
<td>Modified Rankin scale (mRS) (see Supplementary 2 &amp; 3)</td>
<td>2=horizontal stacked bar charts (two bars versus four) 1=table 2=pie charts</td>
<td>1=horizontal stacked bar chart with two bars 2=pie charts with &gt;6 segments</td>
</tr>
<tr>
<td>EuroQol Five Dimensions Visual Analogue Scale (EQ-VAS) (see Supplementary 4)</td>
<td>5=line graphs 1=box plot 1=VAS infographic with normative data 1=caterpillar plot 1=table 2=dot plots</td>
<td>3=line graphs 1=dot plot</td>
</tr>
<tr>
<td>EuroQol Five Dimensions Three Levels (EQ-5D-3L) (see Supplementary 5 &amp; 6)</td>
<td>3=horizontal bar charts (one pair with a table) 1=vertical stacked bar chart 1=spider chart 1=icon array/infographic 1=table</td>
<td>2=horizontal bar charts 1=spider chart 1=icon array/infographic 1=table</td>
</tr>
</tbody>
</table>
Feedback of aggregate, health-service level patient-reported outcomes data

<table>
<thead>
<tr>
<th>Template category</th>
<th>Stage 1: Co-design workshop process*</th>
<th>Stage 2: Online survey</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Workshop 2</td>
<td>Workshop 3</td>
</tr>
<tr>
<td><strong>Formats presented</strong></td>
<td><strong>Formats retired</strong></td>
<td><strong>Formats presented</strong></td>
</tr>
<tr>
<td>Hospital Anxiety Depression Scale</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(HADS)</td>
<td>2=horizontal bar charts</td>
<td>2=vertical bar charts</td>
</tr>
<tr>
<td></td>
<td>1=dual pie chart display</td>
<td>2=vertical bar charts</td>
</tr>
<tr>
<td></td>
<td>1=horizontal bar chart</td>
<td>1=horizontal stacked</td>
</tr>
<tr>
<td></td>
<td></td>
<td>bar chart</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1=dual pie chart display</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2=line charts</td>
</tr>
<tr>
<td>Data over time*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(see Supplementary 8 &amp; 9)</td>
<td>No template shown</td>
<td>No template shown</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No template shown</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No template shown</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Formats included:
- Format C=vertical bar charts with local and peer comparison data
- Format D=horizontal bar charts with state-level hospital comparison data
- Format E=horizontal bar charts with state-level hospital comparison data (paired with a table)
- Format F=horizontal stacked bar charts
- Format G=horizontal stacked bar charts with separated graphs for each dimension of health
- Format A=vertical bar charts with state-level comparison hospital data
- Format B=vertical bar charts with state-level comparison hospital data (including clinical cut-off)
- Format C=horizontal stacked bar chart with state-level comparison hospital data
- Format D=dual pie chart display
Feedback of aggregate, health-service level patient-reported outcomes data

<table>
<thead>
<tr>
<th>Template category</th>
<th>Stage 1: Co-design workshop process*</th>
<th>Stage 2: Online survey</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Workshop 2</td>
<td>Workshop 3</td>
</tr>
<tr>
<td></td>
<td>Formats presented</td>
<td>Formats retired</td>
</tr>
<tr>
<td></td>
<td>Format C=vertical bar charts with local hospital data only</td>
<td>Format D=vertical bar charts with state-level hospital comparison data</td>
</tr>
<tr>
<td>Additional formats</td>
<td>1=forest plot</td>
<td>1=heat map</td>
</tr>
</tbody>
</table>

Table: refers to a numerical table with data values. *Workshop 1 is excluded as no templates were formally assessed during the first introductory workshop.
### Feedback of aggregate, health-service level patient-reported outcomes data

#### Table III. Summary of outcomes according to each project stage

<table>
<thead>
<tr>
<th>Patient Reported Outcome Measure</th>
<th>Stage 1</th>
<th>Stage 2</th>
<th>Stage 3 (Final Consensus)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Modified Rankin Scale (mRS)</strong></td>
<td>• Support for stacked bar chart format included comparative data (i.e. peer- state- and national- hospital level)</td>
<td>• Stacked bar chart with proportions for each of the levels of disability and comparative data for local hospital against peer- and state-hospitals was identified as most preferred</td>
<td>• A stacked bar chart was the preferred format to feedback aggregate mRS data to clinicians. There is a need to consider the needs of the target audience to determine whether additional information should also be included in a table with data values</td>
</tr>
<tr>
<td></td>
<td>• Support for inclusion of table with data values to supplement a stacked bar chart</td>
<td>• Pie chart and doughnut chart were least preferred</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Mixed support for pie charts, with a maximum of three segments recommended</td>
<td>• Non-clinical respondents preferred the inclusion of a table with data values to supplement the stacked bar chart</td>
<td></td>
</tr>
<tr>
<td><strong>EuroQol Five Dimensions Three Levels (EQ-5D-3L)</strong></td>
<td>• Support for bar charts and stacked bar charts</td>
<td>• Stacked bar chart that presented all five dimensions of the EQ-5D-3L and presented each dimension according to the proportion of patients reporting ‘no problems’ versus those reporting ‘problems’ was identified as the preferred format</td>
<td>• A stacked bar chart with colour coding was identified as the prevailing preference for the feedback of aggregate EQ-5D-3L data</td>
</tr>
<tr>
<td></td>
<td>• The inclusion of comparative data in bar and stacked bar chart formats was preferred in comparison to the information presented in a spider chart or pictograph</td>
<td>• Non-clinical respondents preferred the inclusion of a table with data values to supplement the stacked bar chart</td>
<td></td>
</tr>
<tr>
<td><strong>EuroQol Five Dimensions Three Levels Visual Analogue Scale (EQ-VAS)</strong></td>
<td>• Support was identified for a scatterplot</td>
<td>• A line graph was preferred for reporting EQ-VAS scores over time. A separate category for longitudinal data was supported</td>
<td>• A bar chart including confidence intervals was identified as the preferred template for the feedback of EQ-VAS group-level data</td>
</tr>
<tr>
<td></td>
<td>• Support was identified for bar chart</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Feedback of aggregate, health-service level patient-reported outcomes data

<table>
<thead>
<tr>
<th>Patient Reported Outcome Measure</th>
<th>Stage 1</th>
<th>Stage 2</th>
<th>Stage 3 (Final Consensus)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Stage 1</strong></td>
<td>✔️ Opposed the inclusion of a table with data values to supplement a graph/chart format</td>
<td>✔️ Boxplots would need instructions to explain the graph (e.g. define the median, confidence intervals etc.)</td>
<td>✔️ A bar chart including a clinical cut-off point was identified as the preferred format for the feedback of HADS data</td>
</tr>
<tr>
<td><strong>Hospital Anxiety Depression Scale (HADS)</strong></td>
<td>✔️ Support for pie charts, ✔️ Support for bar charts, ✔️ Support for the inclusion of comparative data in bar chart formats</td>
<td>✔️ Bar charts with the inclusion of clinically meaningful cut-off points were preferred by clinicians, ✔️ Pie chart was preferred by non-clinical respondents</td>
<td>✔️ A line graph with comparative data (e.g. peer hospital, state or national-level data) was recommended as the preferred template for the presentation of data over time, ✔️ The inclusion of additional information, such as associations with processes of clinical care were considered too complex for routine feedback purposes</td>
</tr>
<tr>
<td><strong>Display of Patient-Reported Outcomes data over time</strong></td>
<td>✔️ Opposed the stratification of data by patient age group and sex within the one format due to complexity/cognitive burden</td>
<td>✔️ Line graph stratified by patient characteristics was identified as the preferred format, ✔️ Bar chart were least preferred for presenting data over time</td>
<td>✔️ A line graph with comparative data (e.g. peer hospital, state or national-level data) was recommended as the preferred template for the presentation of data over time</td>
</tr>
<tr>
<td><strong>General formatting features</strong></td>
<td>✔️ Use simple layouts with text and symbols to reduce cognitive burden</td>
<td>✔️ Use intuitive colouring (i.e. green=better/good outcomes and red=worse/poor outcomes), while ensuring that colours in templates</td>
<td>✔️ Care must be taken in the comparison of health service benchmarks against normative population values that may be</td>
</tr>
</tbody>
</table>

---

**BMJ Open**

Feedback of aggregate, health-service level patient-reported outcomes data

<table>
<thead>
<tr>
<th>Patient Reported Outcome Measure</th>
<th>Stage 1</th>
<th>Stage 2</th>
<th>Stage 3 (Final Consensus)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• Use definitions and instructions (where appropriate)</td>
<td>• Use clear labelling, including graph axes and titles</td>
<td>• Derived from a different cohort or clinical population</td>
</tr>
<tr>
<td></td>
<td>• Define all acronyms when used</td>
<td>• Include sample sizes and proportions</td>
<td>• Appropriate case-mix adjustment methods for some PROMs need to be developed</td>
</tr>
<tr>
<td></td>
<td>• Include confidence intervals or clinically meaningful differences (where appropriate/available)</td>
<td>• Include comparative data to facilitate benchmarking between health services</td>
<td>• A need for education and training support among clinicians based on the use and interpretation of PROs data was identified and supported</td>
</tr>
<tr>
<td></td>
<td>• Avoid using colours that are difficult to differentiate between</td>
<td>• Include instructions for how to interpret the chart</td>
<td>• Additional pilot testing of the recommended feedback templates may also be required to refine the formats for different target audiences</td>
</tr>
<tr>
<td></td>
<td>• Clearly label exclusions used in the calculations of the results presented in the charts</td>
<td>• Allow spacing between columns of bar charts/Graphs</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Include sample sizes</td>
<td>• Avoid multiple charts on the one page</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Avoid use of spider plots, contour plots and heat maps due to greater level of complexity</td>
<td>• Avoid too much/repetitive text</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Avoid using pictographs due to their simplicity</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

BMJ Publishing Group Limited (BMJ) disclaims all liability and responsibility arising from any reliance Supplemental material placed on this supplemental material which has been supplied by the author(s) BMJ Open

doi: 10.1136/bmjopen-2021-055999

BMJ Open

Feedback of aggregate, health-service level patient-reported outcomes data

### Table IV: Median ratings for the templates tested in the online survey (stage two)

<table>
<thead>
<tr>
<th>Patient-Reported Outcome Measure</th>
<th>Template Reference*</th>
<th>Total survey responses N</th>
<th>Order of overall preference</th>
<th>Median perceived ease of understanding* (Q1, Q3)</th>
<th>Median perceived usefulness to clinical practice* (Q1, Q3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>modified Rankin Scale (mRS)</td>
<td>Format A</td>
<td>95</td>
<td>1</td>
<td>9 (8,10)</td>
<td>8 (7,9)</td>
</tr>
<tr>
<td></td>
<td>Format B</td>
<td>94</td>
<td>2</td>
<td>7 (6,9)</td>
<td>7 (5,8)</td>
</tr>
<tr>
<td></td>
<td>Format C</td>
<td>94</td>
<td>3</td>
<td>9 (7,10)</td>
<td>7 (5,8)</td>
</tr>
<tr>
<td></td>
<td>Format D</td>
<td>91</td>
<td>4</td>
<td>8 (6,9)</td>
<td>7 (5,8)</td>
</tr>
<tr>
<td></td>
<td>Format E</td>
<td>91</td>
<td>5</td>
<td>7 (4,8)</td>
<td>6 (4,8)</td>
</tr>
<tr>
<td>EuroQol Visual Analogue Scale (EQ-VAS)</td>
<td>Format A</td>
<td>82</td>
<td>1</td>
<td>7 (6,8)</td>
<td>7 (5,8)</td>
</tr>
<tr>
<td></td>
<td>Format B</td>
<td>82</td>
<td>2</td>
<td>6 (4,8)</td>
<td>5 (4,8)</td>
</tr>
<tr>
<td></td>
<td>Format C</td>
<td>82</td>
<td>3</td>
<td>6 (4,7)</td>
<td>6 (4,7)</td>
</tr>
<tr>
<td>EuroQol Five Dimension Descriptive System</td>
<td>Format A</td>
<td>78</td>
<td>2</td>
<td>9 (8,10)</td>
<td>8 (7,9)</td>
</tr>
<tr>
<td></td>
<td>Format B</td>
<td>78</td>
<td>4</td>
<td>8 (7,9)</td>
<td>8 (7,9)</td>
</tr>
<tr>
<td></td>
<td>Format C</td>
<td>78</td>
<td>7</td>
<td>6 (4,7)</td>
<td>6 (4,7)</td>
</tr>
<tr>
<td></td>
<td>Format D</td>
<td>78</td>
<td>5</td>
<td>8 (7,9)</td>
<td>8 (6,9)</td>
</tr>
<tr>
<td></td>
<td>Format E</td>
<td>78</td>
<td>3</td>
<td>8 (6,9)</td>
<td>7 (6,9)</td>
</tr>
<tr>
<td></td>
<td><strong>Format F</strong></td>
<td><strong>78</strong></td>
<td><strong>1</strong></td>
<td><strong>7 (6,9)</strong></td>
<td><strong>7 (6,8)</strong></td>
</tr>
<tr>
<td></td>
<td>Format G</td>
<td>78</td>
<td>6</td>
<td>6 (5,8)</td>
<td>6 (5,8)</td>
</tr>
<tr>
<td>Hospital Anxiety and Depression Scale</td>
<td>Format A</td>
<td>75</td>
<td>4</td>
<td>9 (8,10)</td>
<td>8 (7,9)</td>
</tr>
<tr>
<td></td>
<td><strong>Format B</strong></td>
<td><strong>75</strong></td>
<td><strong>1</strong></td>
<td><strong>9 (8,10)</strong></td>
<td><strong>8 (7,9)</strong></td>
</tr>
<tr>
<td></td>
<td>Format C</td>
<td>75</td>
<td>3</td>
<td>8 (7,9)</td>
<td>8 (6,9)</td>
</tr>
<tr>
<td></td>
<td>Format D</td>
<td>75</td>
<td>2</td>
<td>8 (6,9)</td>
<td>7 (5,9)</td>
</tr>
<tr>
<td>Longitudinal data</td>
<td>Format A</td>
<td>69</td>
<td>4</td>
<td>8 (7,9)</td>
<td>7 (6,8)</td>
</tr>
<tr>
<td></td>
<td>Format B</td>
<td>69</td>
<td>2</td>
<td>8 (7,9)</td>
<td>7 (7,9)</td>
</tr>
<tr>
<td></td>
<td>Format C</td>
<td>69</td>
<td>5</td>
<td>8 (7,9)</td>
<td>7 (6,8)</td>
</tr>
<tr>
<td></td>
<td>Format D</td>
<td>69</td>
<td>3</td>
<td>8 (7,9)</td>
<td>7 (6,8)</td>
</tr>
<tr>
<td></td>
<td><strong>Format E</strong></td>
<td><strong>69</strong></td>
<td><strong>1</strong></td>
<td><strong>8 (7,9)</strong></td>
<td><strong>7 (6,8)</strong></td>
</tr>
<tr>
<td></td>
<td>Format F</td>
<td>69</td>
<td>3*</td>
<td>7 (5,8)</td>
<td>6 (5,8)</td>
</tr>
</tbody>
</table>

Bolded rows represent the prevailing preference identified for each Patient Reported Outcome Measure category according to the online survey data. *See Supplemental files 2 to 9 for the templates used for each format. *Based on scale response data where: 0=very difficult, 10=very easy; where 0= not at all, 10=very useful. #Equal third preference.
### Feedback of aggregate, health-service level patient-reported outcomes data

**Table V: Recommendations to guide feedback of aggregate, summary Patient Reported Outcomes data to hospital end-users**

#### General principles

1. Use bar charts and stacked bar charts when reporting aggregate PROs data at a single time point
2. Use line graphs when reporting aggregated PROs data longitudinally
3. Bar charts should include comparative data (e.g. from other health services, state or national data)
4. Avoid complex graphs such as spider plots, contour plots and heat maps
5. Avoid pictographs for providing summary PROs data feedback to clinicians
6. Use tools to assist with interpretation including instructions, labelling of exclusions and sample sizes
7. Include achievable benchmarks and comparative hospital data ensuring clear definitions of benchmarks or comparators are provided
8. Careful consideration must be given for the inclusion of hospital achievable benchmarks versus normative population values
9. Aggregated PROs data should ideally be fed back to hospital staff at 3-6-month intervals
10. Involve hospital end-users if aggregate PROs summary data are being prepared by academics/researchers
11. Present PROs summary data in more than one way to address the needs of specific audience sub-groups (e.g. clinical versus non-clinical recipients/medical, nursing or allied health professionals etc.)
12. Further methodological research to determine the most appropriate analytic methods and the variables needed to enable reliable case-mix adjustment PROs summary data is required

#### General presentation and formatting considerations

13. Allow space between the columns of bar graphs, and include details such as p-values, confidence intervals, clinically important differences, sample sizes and the proportions within the graph (as relevant)
14. Pie charts should have a maximum of two to three segments
15. Clearly label graphs, including the axes and a title
16. Avoid acronyms within a graph or ensure these are defined when used
17. Ensure that graphs do not look visually ‘cluttered’
18. Use intuitive colouring to depict data in graphical formats (e.g. green for positive outcomes and red for negative outcomes)
19. Ensure that colours in templates are easily discernible when printed in grayscale and avoid the use of similar colours

#### Specific considerations for reporting of Quality of Life outcomes using the EQ-5D-3L as one of the most common generic PROMs used by CQRs

20. Use stacked bar charts to report all five dimensions of EQ-5D-3L in a single template
21. Incorporate comparative hospital data (if available)
22. Consider classifying responses to two categories: patients reporting ‘no problems’ as differentiated from those reporting ‘problems’

23. Use intuitive colouring to depict data in graphical formats (e.g. green for ‘no problems’)

PROs: Patient-reported outcomes. EQ-5D-3L: EuroQol Five Dimensions Three Levels.
Feedback of aggregate, health-service level patient-reported outcomes data

Supplemental Acknowledgments: Patient Reported Outcome Measures Working Group and other contributors

The following people are acknowledged for their contribution to the Patient Reported Outcome Measures Pilot Project which this manuscript is based on:

Project Team

Jennifer Tragardh (Victorian Stroke Clinical Network); Joosup Kim (Monash University); Julie Morrison (AuSCR/The Florey Institute of Neuroscience and Mental Health); Natasha Lannin (Alfred Health/Monash University); Tara Purvis (Monash University); Tharshanah Thayabaranathan (Monash University); Sharon Kramer (The Florey Institute of Neuroscience and Mental Health).

Project Working Group Members

We acknowledge the following members for their input and expertise in developing and refining the PROs data templates:

Ben Clissold (Barwon Health); Bronwyn Coulton (Monash Health); Casey Hair (Ballarat Health Services); Christy Hatherley (Mental Health Clinical Network/Safer Care Victoria); Claire Formby (St Vincent’s Hospital, Melbourne); Colin Scott (Consumer Representative); Dimitra Chrisikakos (St George’s Hospital Community Rehabilitation Centre); Geoff Cloud (Alfred Health); Jo Wrench (Austin Health); Julia Brown (St Vincent’s Hospital, Melbourne); Julian Ellis (Alfred Health); Kanaga Lagma (Peninsula Health – Frankston); Lauren Guy (Western Health); Lauren Sanders (St Vincent’s Hospital, Melbourne); Lisa Mangwiro (Stroke Connect Victoria – Stroke Foundation); Marissa Stone (St Vincent’s Hospital, Melbourne); Patrick Groot (Warrnambool Base Hospital); Penina Gunzburg (Alfred Health); Peter Hand (Alfred Health); Robyn Soulsby (Bendigo Health); Shae Cooke (Eastern Health); Tessa Coupland (Bendigo Health).

Clinical and Academic Networks

We are most appreciative to the following organizations and clinical/academic networks that assisted us in distributed the link to participate in the online survey used in Stage II of the project:

Australian Stroke Coalition (ASC); Stroke Society of Australasia (SSA); Victorian Stroke Clinical Network (VSCN), including INSIGHT Committee; All Victorian Clinical Networks via Safer Care Victoria (SCV) Clinical eNews; Centre for Research Excellence (CRE) in Stroke Rehabilitation and Brain Recovery; Australian Stroke Clinical Registry (AuSCR) – participating hospitals/clinicians; Department of Health and Human Services Cancer Strategy & Development Unit; Monash University Clinical Registry Special Interest Group; NSW and the ACT stroke clinicians via Stroke Foundation.
Feedback of aggregate, health-service level patient-reported outcomes data

Supplementary Files

File 2: modified Rankin Scale templates (Format A to D)
File 3: modified Rankin Scale templates (continued) (Format E)
File 4: EQ-VAS templates (Format A to C)
File 5: EQ-5D-3L descriptive system data templates (Format A to D)
File 6: EQ-5D-3L descriptive system data templates continued (Format E to G)
File 7: HADS templates (Format A to D)
File 8: Longitudinal data templates (Format A to D)
File 9: Longitudinal data templates (Format E to F)
Modified Rankin Scale templates (Format A to E)

**Format A:**

Self-reported Modified Rankin Scale (mRS) scores at 90-180 days after admission

<table>
<thead>
<tr>
<th>Format</th>
<th>Hospital</th>
<th>My Hospital (N=143)</th>
<th>All Victorian Hospitals (N=4777)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0: No symptoms</td>
<td>23%</td>
<td>20%</td>
<td>17%</td>
</tr>
<tr>
<td>1: No Significant Disability</td>
<td>24%</td>
<td>20%</td>
<td>17%</td>
</tr>
<tr>
<td>2: Significant Disability</td>
<td>22%</td>
<td>20%</td>
<td>17%</td>
</tr>
<tr>
<td>3: Moderately Severe Disability</td>
<td>17%</td>
<td>18%</td>
<td>18%</td>
</tr>
<tr>
<td>4: Severe Disability</td>
<td>6%</td>
<td>5%</td>
<td>6%</td>
</tr>
<tr>
<td>5: Extreme Disability</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
</tbody>
</table>

0: No symptoms
1: No Significant Disability
2: Moderate Disability
3: Severe Disability
4: Moderate Severe Disability
5: Extreme Disability
6: Dead

**Instructions:**

The modified Rankin Scale (mRS) is a measure of functional status recorded on an ordinal scale with 6 categories ranging from zero (no symptoms at all) to five (complete physical dependence). The sixth category can be added to signify death.

Report for July 01 2019 to December 31 2018 admissions (excluding in-hospital deaths).

**Format B:**

Self-reported Modified Rankin Scale (mRS) scores at 90-180 days after admission

**Format C:**

Self-reported Modified Rankin Scale (mRS) scores at 90-180 days after admission

<table>
<thead>
<tr>
<th>Format</th>
<th>Hospital</th>
<th>My Hospital (N=143)</th>
<th>All Victorian Hospitals (N=4777)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0: No symptoms</td>
<td>41%</td>
<td>39%</td>
<td>61%</td>
</tr>
<tr>
<td>1: No Significant Disability</td>
<td>59%</td>
<td>5%</td>
<td>1%</td>
</tr>
<tr>
<td>2: Significant Disability</td>
<td>3%</td>
<td>6%</td>
<td>1%</td>
</tr>
<tr>
<td>3: Moderately Severe Disability</td>
<td>2%</td>
<td>4%</td>
<td>1%</td>
</tr>
<tr>
<td>4: Severe Disability</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>5: Extreme Disability</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
</tbody>
</table>

0: No symptoms
1: No Significant Disability
2: Moderate Disability
3: Severe Disability
4: Moderate Severe Disability
5: Extreme Disability
6: Dead

**Instructions:**

The modified Rankin Scale (mRS) is a measure of functional status recorded on an ordinal scale with 6 categories ranging from zero (no symptoms at all) to five (complete physical dependence). The sixth category can be added to signify death.

Report for July 01 2019 to December 31 2018 admissions (excluding in-hospital deaths).

**Format D:**

Self-reported Modified Rankin Scale (mRS) scores at 90-180 days after admission

**Format E:**

Self-reported Modified Rankin Scale (mRS) scores at 90-180 days after admission

**Instructions:**

The modified Rankin Scale (mRS) is a measure of functional status recorded on an ordinal scale with 6 categories ranging from zero (no symptoms at all) to five (complete physical dependence). The sixth category can be added to signify death.

Report for July 01 2019 to December 31 2018 admissions (excluding in-hospital deaths).
Modified Rankin Scale templates (Format A to E) (continued)

Format E:

Self-reported Modified Rankin Scale (mRS) scores at 90-180 days after admission

- 61% Dependents with marked disability (mRS 3-5)
- 41% Independent patients (mRS 0-2)

My Hospital (N=43)

All Victorian Hospitals (N=477)

Instructions:
The modified Rankin scale (mRS) is a measure of functional status recorded on a ordinal scale of six categories ranging from no problems (mRS 0) to fatal performance (mRS 6) (e.g. unable to walk). The report displays the proportion of independent patients (mRS 0-2) compared to the proportion of dependent patients with marked disability (mRS 3-5).

The bubble chart that presented in this report displays the proportion of patients from your hospital in the centre circle compared to all Victorian Hospitals in the outer circle.

Report for July 01 2018 to December 31 2018 admissions (excluding in hospital deaths).
EuroQol Visual Analogue Scale templates (Format A to C)

Format A:

Mean Visual Analogue Scale (VAS) results at 90 to 180 days after admission from YOUR HOSPITAL

Format B:

Median Visual Analogue Scale (VAS) results at 90 to 180 days after admission

Format C:

Median Visual Analogue Scale (VAS) results at 90 to 180 days after admission
EuroQol Five Dimension Descriptive System templates (Format A to G)

**Format A:**
EuroQol Five Dimension (EQ-SD) results to anxiety/depression at 90 to 180 days after admission at YOUR HOSPITAL

![Image showing EQ-SD results]

**Instructions:**
The EuroQol Five Dimension (EQ-SD) is a measure of health-related quality of life, with five dimensions covering the following areas:
- Mobility
- Self-care
- Usual activities
- Pain/discomfort
- Anxiety/depression

Each dimension has three possible responses: no problems, some problems, and severe problems.

The stacked bar chart is a display of the proportion of patients reporting to one of the dimensions, anxiety/depression, at your hospital.

Report for July 01, 2021 to December 31, 2018 admissions.

**Format B:**
EuroQol Five Dimension (EQ-SD) results to anxiety/depression at 90 to 180 days after admission by age group

![Image showing EQ-SD results by age group]

**Instructions:**
The EuroQol Five Dimension (EQ-SD) is a measure of health-related quality of life, with five dimensions covering the following areas:
- Mobility
- Self-care
- Usual activities
- Pain/discomfort
- Anxiety/depression

Each dimension has three possible responses: no problems, some problems, and severe problems.

The stacked bar chart is a display of the proportion of patients reporting to one of the dimensions, anxiety/depression, at your hospital.

Report for July 01, 2021 to December 31, 2018 admissions.

**Format C:**
EQ-SD (Quality of Life) Metrics:
Self-reported anxiety/depression at 90-180 days after admission, by age group

![Image showing EQ-SD metrics]

**Instructions:**
The EuroQol Five Dimension (EQ-SD) is a measure of health-related quality of life, with five dimensions covering the following areas:
- Mobility
- Self-care
- Usual activities
- Pain/discomfort
- Anxiety/depression

Each dimension has three possible responses: no problems, some problems, and severe problems.

The stacked bar chart is a display of the proportion of patients reporting to one of the dimensions, anxiety/depression, at your hospital and at all Victorian Hospitals.

Report for July 01, 2018 to December 31, 2018 admissions.

**Format D:**
Proportion of patients reporting problems across dimensions of the EQ-SD-3L at 90-180 days after admission

![Image showing proportion of patients reporting problems]

**Instructions:**
The EQ-SD-3L includes 5 dimensions shown on the y-axis (mobility, self-care, usual activities, pain/discomfort, anxiety/depression).

Each dimension consists of 5 levels of problems (level 0: no problems, level 1: some problems, level 2: moderate problems, level 3: severe problems). The figure shows how many patients report problems in each dimension.

Report for July 01, 2018 to December 31, 2018 admissions.
EuroQol Five Dimension Descriptive System templates (Format A to G)

Format E:
Proportion of patients reporting problems across dimensions of the EQ-5D-3L at 90-180 days after admission

Instructions:
The EQ-5D-3L includes 5 dimensions shown on the y-axis (mobility, self-care, usual activities, pain/discomfort, anxiety/depression). Each dimension consists of 3 levels no problems (Level 1), some problems (Level 2) and extreme problems (Level 3). The figures shown here includes Level 2 and 3 (i.e. patients reporting problems for each dimension).

Format G:
Self-reported EQ-5D-3L results at 90-180 days after admission AT YOUR HOSPITAL

Instructions:
The EQ-5D-3L includes 5 dimensions (mobility, self-care, usual activities, pain/discomfort, anxiety/depression). Each dimension consists of 3 levels no problems (Level 1), some problems (Level 2) and extreme problems (Level 3). The green bars display the proportion of patients reporting no problems (Level 1) for each dimension, and the red bar displays the proportion of patients reporting problems (Level 2 and 3).
Templates presenting longitudinal data (Format A to F)

**Format A:**

Mean Visual Analogue Scale (VAS) results at 90 to 180 days after admission

![Graph showing longitudinal data for Format A]

**Instructions:**

- **Dotted line** indicates expected outcomes based on normative population aged 70-79.
- **Blue line** indicates your hospital’s actual VAS outcomes over time.
- Confidence intervals indicate the range in which we are 95% confident that the true mean lies within the intervals.
- The number of patients who completed the follow-up interview each year.
- The Visual Analogue Scale (VAS) is a measure of health-related quality of life. The VAS records the patients’ self-rated health on a vertical, visual analogue scale where the endpoints are labelled “worst imaginable health state” (0) and “best imaginable health state” (100).

**Mean Visual Analogue Scale (VAS) Score**

<table>
<thead>
<tr>
<th>Year</th>
<th>2016</th>
<th>2017</th>
<th>2018</th>
<th>2019</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score</td>
<td>92.8</td>
<td>89.7</td>
<td>91.5</td>
<td>89.3</td>
</tr>
</tbody>
</table>

**Format B:**

Mean Visual Analogue Scale (VAS) results at 90 to 180 days after admission

![Graph showing longitudinal data for Format B]

**Instructions:**

- **Dotted line** indicates expected outcomes based on normative population aged 70-79.
- **Blue line** indicates your hospital’s actual VAS outcomes over time.
- Confidence intervals indicate the range in which we are 95% confident that the true mean lies within the intervals.
- The number of patients who completed the follow-up interview each year.
- The Visual Analogue Scale (VAS) is a measure of health-related quality of life. The VAS records the patients’ self-rated health on a vertical, visual analogue scale where the endpoints are labelled “worst imaginable health state” (0) and “best imaginable health state” (100).

**Mean Visual Analogue Scale (VAS) Score**

<table>
<thead>
<tr>
<th>Year</th>
<th>2016</th>
<th>2017</th>
<th>2018</th>
<th>2019</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score</td>
<td>91.2</td>
<td>89.8</td>
<td>91.0</td>
<td>89.5</td>
</tr>
</tbody>
</table>

**Format C:**

Mean Visual Analogue Scale (VAS) results at 90 to 180 days after admission

![Graph showing longitudinal data for Format C]

**Instructions:**

- **Dotted line** indicates expected outcomes based on normative population aged 70-79.
- **Blue line** indicates your hospital’s actual VAS outcomes over time.
- Confidence intervals indicate the range in which we are 95% confident that the true mean lies within the intervals.
- The number of patients who completed the follow-up interview each year.
- The Visual Analogue Scale (VAS) is a measure of health-related quality of life. The VAS records the patients’ self-rated health on a vertical, visual analogue scale where the endpoints are labelled “worst imaginable health state” (0) and “best imaginable health state” (100).

**Mean Visual Analogue Scale (VAS) Score**

<table>
<thead>
<tr>
<th>Year</th>
<th>2016</th>
<th>2017</th>
<th>2018</th>
<th>2019</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score</td>
<td>92.5</td>
<td>90.3</td>
<td>91.8</td>
<td>89.6</td>
</tr>
</tbody>
</table>

**Format D:**

Mean Visual Analogue Scale (VAS) results at 90 to 180 days after admission

![Graph showing longitudinal data for Format D]

**Instructions:**

- **Dotted line** indicates expected outcomes based on normative population aged 70-79.
- **Blue line** indicates your hospital’s actual VAS outcomes over time.
- Confidence intervals indicate the range in which we are 95% confident that the true mean lies within the intervals.
- The number of patients who completed the follow-up interview each year.
- The Visual Analogue Scale (VAS) is a measure of health-related quality of life. The VAS records the patients’ self-rated health on a vertical, visual analogue scale where the endpoints are labelled “worst imaginable health state” (0) and “best imaginable health state” (100).

**Mean Visual Analogue Scale (VAS) Score**

<table>
<thead>
<tr>
<th>Year</th>
<th>2016</th>
<th>2017</th>
<th>2018</th>
<th>2019</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score</td>
<td>91.0</td>
<td>89.7</td>
<td>91.0</td>
<td>89.5</td>
</tr>
</tbody>
</table>
Templates presenting longitudinal data (Format A to F)

**Format E:**

Mean Visual Analogue Scale (VAS) results at 90 to 180 days after stroke admission

- **Instructions:**
  - Dotted line indicates expected outcomes based on normative population aged 70 to 79 years
  - Dashed line indicates ischemic stroke patient VAS outcomes at YOUR HOSPITAL
  - Solid blue line indicates hemorrhage (ICH) patient VAS outcomes at YOUR HOSPITAL
  - Solid orange line indicates intracerebral hemorrhage (ICH) patient VAS outcomes at ALL VICTORIAN HOSPITALS
  - Red bars indicate number of patients who completed the follow-up interview each year

The Visual Analogue Scale (VAS) is a measure of health-related quality of life. The measure records the patients’ health state on a horizontal line where the endpoints are labeled “Best imaginable health state” (100) and “Worst imaginable health state” (0).

**Format F:**

Mean Visual Analogue Scale (VAS) results at 90 to 180 days after admission

- Instructions:
  - Dotted line indicates expected outcomes based on normative population aged 70-79 years
  - Blue bars indicate ischemic stroke patient VAS outcomes at YOUR HOSPITAL
  - Light blue bars indicate ischemic stroke patient VAS outcomes at ALL VICTORIAN HOSPITALS
  - Orange bars indicate intracerebral hemorrhage (ICH) patient VAS outcomes at YOUR HOSPITAL
  - Light orange bars indicate intracerebral hemorrhage (ICH) patient VAS outcomes at ALL VICTORIAN HOSPITALS
  - Red bars indicate number of patients who completed the follow-up interview each year

The Visual Analogue Scale (VAS) is a measure of health-related quality of life. The measure records the patients’ self-rated health on a horizontal line, where the endpoints are labeled “Best imaginable health state” (100) and “Worst imaginable health state” (0).