Patient-reported outcome measures (PROMs) in Victoria: who, why, what, when?

Collective views from the PROMs workshop held on 21 October 2016.

Attendees: An Duy Tran, Anthony Harris, Arthur Hsueh, Bernice Ma, Cathy Mihalopoulos, Chris Schilling, Duncan Mortimer, Hannah Jackson, Jenny Watts, Jongsay Yong, Josh Knight, Kim Dalziel, Li Huang, Lidia Engel, Nicole Black, Peter Sivey, Philip Clarke, Rachel Knott, Utsana Tonmukayakul, Xinyang Hua from Deakin University, La Trobe University, Monash University, RMIT University, and University of Melbourne.

For further discussion, contact Professor Philip Clarke, http://mspgh.unimelb.edu.au/centres-institutes/centre-for-health-policy/research-group/health-economics
Contents

Patient-reported outcome measures (PROMs) in Victoria: who, why, what, when?............................... 1

Who should we be collecting PROMs from? .......................................................................................... 4

Who’s going to collect it? .......................................................................................................................... 5

Why: why collecting PROMs .................................................................................................................. 6

1. Hospital performance comparisons ..................................................................................................... 6

2. Patient safety improvement and public reporting .............................................................................. 6

3. Individual-level feedback to doctors and patients .......................................................................... 7

4. Research investment: access to data and governance ...................................................................... 7

5. Understand inequalities .................................................................................................................... 8

6. Pay-for-performance ........................................................................................................................ 8

What: what should be collected? ............................................................................................................. 11

1. Principles for selecting PROMs ......................................................................................................... 11

2. Response to proposals re selecting PROMS included in DHHS consultation paper .................. 12

3. Recommendations .......................................................................................................................... 15

When: When to measure PROMs? ........................................................................................................ 21

Background ........................................................................................................................................... 21

Recommendations ................................................................................................................................. 22
Who should we be collecting PROMs from?

**Pilot study**

Consider the following criteria when making choice of population for pilot:

- Favour high prevalence, severe disease and/or high cost patients and procedures
- Favour clinical areas where we know there are effective interventions or cures available
  
  **-Focus on patients where there is a high likelihood of observing significant improvements in quality of life**

- Focus on where there is variation in the state in terms of practice and likely outcomes
  
  **-Focus on a small group of conditions that meet above criteria**

- Consider diversity of patient characteristics e.g. cultural, age, SES in order to maximise learnings
  
  **-Measure selected conditions across a selected representative sample of Victorian public hospitals so that the results are reasonably comparable**

**Ongoing PROMs expansion**

Following the pilot there will be opportunity to expand PROMs collection. The following are several options for expansion:

- Add additional conditions that meet the same criteria used in pilot
- A broader range of conditions extending to those where there is a lower likelihood of substantial improvement e.g. chronic conditions and more variation in patient capacity to benefit
- PROMs collection for different specific patient groups such as children where there is a requirement for alternative instruments
- Consider expansion to private hospitals
- Consider health services beyond hospital setting
- Measurement of patients on elective surgery waiting lists
Who’s going to collect it?

- We recommend data collection be built into routine hospital administration processes. Consideration of economic implications of data collection should be made with exploration of online and app options.
- If PROMS are expanded beyond the hospital system, there needs to be a health service responsible for this collection.
Why: why collecting PROMs

1. Hospital performance comparisons

Using PROMs to measure hospital performance is a key motivation for collection of PROMs data. Cutting-edge research from the UK has used PROMs to compare providers’ risk-adjusted patient outcomes (Nuttal et al 2015), to measure the trade-off between hospital efficiency and patient outcomes (Gutaker et al 2013), the relationship between waiting times and patient outcomes (Nikolova et al 2016) and how patient outcomes affect choice of hospital (Gutaker et al 2016). Similar research could provide highly policy-relevant information to Victorian policymakers, inform patient choice and improve transparency around hospital performance.

**Recommendations:**

- Make PROMs available linked to hospital administrative datasets such as the VAED and available to researchers with appropriate safeguards for privacy through ethics applications and data security measures
- Focus on PROMS for elective procedures
- PROMs must be consistent (ie the same generic instrument used) across hospitals and conditions

2. Patient safety improvement and public reporting

The recent Duckett report (Duckett et al 2016) has made several recommendations regarding improvement patient safety and quality. Recommendations 4 and 10 from the report mention improving the flow of data, establishing a new Victorian Health Performance Authority and adopting improvement of patients’ engagement and experience as a priority. While not specifically mentioning PROMS, public availability of PROMS measures seems a natural implication of these recommendations. Provider-level data on risk-adjusted PROMs are already being published in the UK ([http://content.digital.nhs.uk/PROMs](http://content.digital.nhs.uk/PROMs)).
3. Individual-level feedback to doctors and patients

PROMs can be used for various forms of individual-level feedback to improve treatment decisions made by patients and health professionals. One is the use of population-level data to inform decision making (e.g., female patients aged 50-60 with BMI 30-40 usually have an improvement in quality of life equal to X from a hip replacement operation). More advanced individual feedback could be using the patient’s real-time PROM responses to inform their clinical decisions including discharge planning and waiting list management (see e.g. Croudace et al. 2016).

**Recommendation**

- Investigate using population PROMs data for health professionals and patients to use for clinical decision making.

- Conduct trials of use of real-time PROMs outcomes in treatment decisions.

4. Research investment: access to data and governance

**Recommendation:**

- We would like to see a commitment to the PROMs data being accessible for research purposes including within higher education and other research institutions. This would include health services research, clinical trials, population health and health economics research. As part of this commitment, the guidelines for access to data must be developed early in the process in tandem with PROMs data collection. The widest possible dissemination of this data while maintaining patient confidentiality will ensure the greatest possible benefit to the community of this data collection.
Collection of PROMs data should be part of a broader suite of investments in analysis of the data on health services research in Victoria

The governance structure of the PROMs data collection must include health services researchers and health economists as key stakeholders and users of the data.

5. Understand inequalities

Studies on socio-economic inequalities in health have usually used simple one-dimensional self-assessed health measures, or survival/mortality rates. However, some studies have begun to use more detailed multi-dimensional measures (eg Burstrom et al 2005). There is considerable scope to gain a much more complete understanding of socio-economic inequalities in health with detailed multi-dimensional health measures for specific conditions such as those made available through PROMs.

**Recommendation**

- Explore opportunities to link PROMs data to socio-economic status as much as possible, for example through patient postcode information.
- Collect socio-economic status variables (income, education) in the PROMs survey itself if possible.

6. Pay-for-performance

Most economists would advocate the principle of aligning reimbursement systems with patient outcomes. However practical applications of pay-for-performance using PROMs in a hospital setting are rare. Recently in the UK, Best Practice Tariffs provide bonuses for hospitals conditional on patient outcomes that do not fall three standard deviations below the mean (Gomes et al 2014), but this only gives incentives to avoid very poor performance, and does not reward better-than average performance.
**Recommendation:**

- Consider trialling limited pay-for-performance schemes using PROMs to inform future policy development.

**References**


What: what should be collected?

1. Principles for selecting PROMs

- PROMs should be fit for purpose (with respect to target and construct psychometric properties, scoring system, mode of administration, cost per administration, and other characteristics of the instrument); a battery of PROMs (with minimal overlap between PROMs) will be required if PROMs data are to be put to many purposes (Sansoni, 2016; ACI, 2016).
  - For the purposes of patient monitoring and practice improvement, PROMs should be sensitive to changes in the condition or disease that would mandate changes in patient management (ACI, 2016 p26).
  - For the purposes of comparisons within conditions or diseases, PROMs should be sensitive to important changes in the condition or disease, provide a summary measure of changes in the condition or disease, and be able to differentiate between important subgroups within the condition- or disease-population (ACI, 2016 p28).
  - For the purposes of comparisons across conditions or diseases, PROMs should provide a ‘common denominator’ capable of capturing important changes in the relevant outcome for a broad range of conditions/diseases (ACI, 2016).
  - For performance measurement and benchmarking at provider- and system-levels, PROMs should be “relevant to stakeholders” (ACI, 2016 p29) and “amenable to change” (ACI, 2016 p19 citing NQF, 2013).
  - For the purposes of informing patient choice and facilitating shared care, PROMs should be “meaningful to the target population” (ACI, 2016 p19 citing NQF, 2013).
- PROMs should be selected to minimise systematic missingness and/or measurement error (Gomes et al, 2016):
  - PROMs should be selected giving due consideration to the trade-off between respondent burden and breadth/depth of information obtained (AIC, 2016 p30).
  - PROMs should be “suitable to be used across different population subgroups… [within] the target population (vulnerable populations, low literacy, language and culture, functional abilities)” (ACI, 2016 p9 citing NQF, 2013).
  - PROMs should be selected giving due consideration to motivation of clinicians to collect the data.
PROMs data are of limited value in isolation (ACI, 2016; Sansoni, 2016; Smith & Street, 2013):

- PROMs data should be linked to information regarding patient characteristics (including risk factors) to permit risk adjustment (ACI, 2016; Sansoni, 2016; Smith & Street, 2013).
- PROMs data should include patient self-report on OOP health care utilisation and should be linked to comprehensive information regarding health service utilisation to facilitate economic evaluation and performance measurement (Sansoni, 2016).
- PROMs data should include patient self-reported condition and be linked to clinical diagnostic information, clinician assessed outcome and other non-patient reported outcomes.

2. Response to proposals re selecting PROMS included in DHHS consultation paper

The DHHS consultation paper (DHHS, 2016) includes a number of specific proposals regarding selection of PROMs for inclusion in the Victorian pilot. These proposals are summarised below and then assessed against the principles for selecting PROMs specified above:

**DHHS Proposal**: “The department has identified the PROMIS-10… as [a] candidate instrument. PROMIS-10 is a computer adaptive testing survey instrument that has the potential to yield more accurate estimates of a patient’s general health status than traditional paper-based instruments” (DHHS, 2016 p7).

**Response**: Selection of the computer adaptive PROMIS-10 may not be consistent with the principles for selecting PROMs specified above:

(i) The [PROMIS-10](#) may not provide a ‘common denominator’ suitable for comparing across conditions or diseases. The PROMIS-10 is comprised of 10 items measuring perceptions of physical health, mental health, social health, pain, fatigue, and overall quality of life. Two issues may limit the extent to which the PROMIS-10 can provide a ‘common denominator’. First, the PROMIS-10 measures relatively coarse-grained variations in physical and mental function and so may struggle to detect important variation in outcomes for some diseases or conditions (e.g. sexual function, sleep disorders and cognitive function). Second, multi-dimensional profile measures such as...
the PROMIS-10, SF12 and SF36 are not designed to generate a summary index score and so further information is required to compare interventions and providers that do better on some items but worse on others. Since development of the PROMIS-10, some progress has been made towards remedying this problem. Summary scales for Global Physical Health (GPH) and Global Mental Health (GMH) can be derived from subsets of PROMIS-10 items using algorithms developed by Hays et al (2009) but higher GPH must still be traded against lower GMH (or vice versa) when comparing non-dominated alternatives. Revicki et al (2009) have estimated cross-walks from PROMIS-10 item scores and summary scores to EQ5D index scores but this constitutes a second-best means of obtaining EQ5D scores (Mortimer & Segal, 2008). A US value set is available for the PROMIS-29 (Craig et al, 2016), permitting derivation of QALYs from PROMIS-29 response data. However, no similar value set is available for the PROMIS-10 and US value sets may not reflect preferences of the Australian population. Derivation of an Australian value set for the PROMIS-10 would provide the ‘common denominator’ necessary for comparisons across conditions and diseases.

(ii) Selection of the computer adaptive version of the PROMIS-10 may not be consistent with minimising systematic missingness and/or measurement error. In the computer adaptive version of the PROMIS-10, respondents “…are only given the minimum number of items that are necessary …to determine their final score and thus respondent burden is far less likely” (Sansoni, 2016 p27). While some studies have reported an association between respondent burden and response rates (Rolstad et al, 2011), other studies have found that response rates, completion rates and results vary depending upon mode of administration (Suris et al, 2007; Hauer et al, 2010; Horevoorts et al, 2015) and that mode of administration has different effects in different population subgroups (Horevoorts et al, 2015). Sansoni (2016) argues that “using online and electronic technologies with all patients may not be possible …and for some patients, paper and pencil forms will still be required” (p35). Where missingness or measurement error varies according to some unobserved patient characteristic that is related to choice of provider or treatment and to health outcomes (e.g. health literacy), there exists the potential for missingness / measurement error to bias comparisons between providers and/or treatments (Gomes et al, 2016).

**DHHS Proposal:** “The department has identified the ICHOM standards… as [a] candidate instrument. The ICHOM standards provide detailed condition-specific assessments and are
potentially comparable across jurisdictions and countries. A number of health services and clinical registries have already begun to collect data or are planning to collect data according to the ICHOM standards”. (DHHS, 2016 p7).

**Response**: ICHOM standards are available for each of the diseases / conditions to be included in the Victorian pilot.¹ Selection of the ICHOM standards may not be consistent with several of the principles for selecting PROMs specified above:

(i) Specifying the ICHOM standards may not achieve the best tradeoff between respondent burden and breadth/depth of information obtained. Each standard includes a battery of PROMs, potentially imposing a significant respondent burden upon patients. For example, the ICHOM standard for older persons includes patient- or carer-reported measures of comorbidities, ADL (Barthel Index, SF36 and gait speed), total number of medications prescribed, total number of adverse drug events and whether medications make the patient feel unwell, hearing/vision impairment, level of care received, falls and fall-related fractures, loneliness and isolation (UCLA 3-item scale), pain (SF36), mood and emotional health (SF36), autonomy and control (the Adult Social Care Outcomes Toolkit), carer burden (4-item Zarit Burden Interview, carer-reported), and preferred place of death.

(ii) The ICHOM standards may not provide a summary measure, sensitive to important changes in the condition or disease. Some ICHOM standards include measures that can be used to summarize changes in the condition or disease, but summary measures for some diseases / conditions are generic measures such as the SF36 (older persons), SF12 (hip and knee OA), and EQ5D (hip and knee OA) rather than condition-specific measures. In recognition of the insensitivity of generic measures to some dimensions of outcome, the ICHOM standards for older persons and hip/knee OA include additional complementary measures of loneliness and isolation (older persons), autonomy and control (older persons), and hip and knee pain (hip/knee OA). Other ICHOM standards include a range of measures to capture changes across different dimensions of patient outcomes (heart failure and localized prostate cancer). In each case, the lack of a sensitive summary measure hampers comparisons between non-dominated alternatives; even within conditions/diseases.

**DHHS Proposal:** “The collection of PROMS data will be accompanied by a collection of complementary data that are needed to contextualise the data. Examples of relevant data include the patient’s age, gender, comorbidities and relevant medical history, and other non-patient reported outcome measures” (p9).

**Response:** While the complementary information nominated in the DHHS consultation paper will allow adjustment for a subset of risk factors. Other relevant factors include BMI, income, education, lifestyle (smoking, exercise, diet), ethnicity, ATSI status, CALD status. To facilitate economic evaluation, linkage to comprehensive data regarding health service utilisation is also required. To facilitate, practice improvement (including monitoring of appropriateness, linkage to data regarding clinical diagnosis and clinician assessed outcomes is also required.

3. Recommendations

**Recommendation 1: Specifying a generic instrument**

1a. Use of the PROMIS-10 has a number of disadvantages such as: sparse coverage of HRQoL space, lack of a summary score, lack of Australian population norms, lack of an Australian value set that would permit direct calculation of QALYs and support economic evaluation.

If PROMIS-10 is to be included as the generic measure in the PROMS pilot, then another measure should be included alongside the PROMIS-10 to address the short-comings of the PROMIS-10, to permit further validation of the PROMIS-10, and to permit derivation of cross-walks to other measures that would support economic evaluation.

Instruments that support economic evaluation, with Australian population norms, and for which there exists an Australian value set for direct estimation of QALYs include: EQ5D, SF36-based SF6D and AQoL-8D. Some of these measures also have better coverage of the HRQoL space than the PROMIS-10 (Richardson et al, 2014).

These alternative measures (AQoL-8D, EQ5D, and SF36-based SF6D) differ with regards to instrument length and patient burden, cost per administration, and coverage of the HRQoL space (Richardson et al, 2015 Table 1; Bryan et al, 2014 Figs 1 & 2, Tables 2 & 3).
**Recommendation 2: Disease- & condition-specific measures**

The addition of disease or condition specific measures should achieve ends that cannot be served by use of the generic instrument. These ends may include: informing patient choice and shared decision making; patient monitoring and practice improvement; and quality assurance, performance measurement and benchmarking.

If existing data collections meet the criteria outlined above for selecting disease- & condition-specific measures, then feasibility, cost and buy-in from clinicians and providers engaged in existing PROMs data collections should be given due weight in any decision to deviate from measures included in those existing data collections.

If ICHOM standards are to be used as the starting point for selection of disease and condition-specific measures, measures should be selected to minimise overlap with generic measures and giving due consideration to the trade-off between patient burden, cost and value of information.

The trade-off between patient burden, cost and value of information should be evaluated during the pilot, ideally by randomising patients to long and short form versions of the data collection instrument.

**Recommendation 3: Special populations**

Special consideration should be given to selecting generic and disease-specific measures that would support economic evaluation for populations in whom standard measures and standard modes of administration may be inappropriate. Where PROMs data collection includes conditions affecting children, the CHU9D is suited to use in children and supports derivation of QALYs for use in economic evaluation (Stevens, 2012). Value sets that would support derivation of QALYs from other child-specific measures are currently under development. Where PROMs data collection encompasses patients with limited capacity, PROMs should be suitable for proxy completion.

**Recommendation 4: Additional information**

4a. PROMs data should be linked to comprehensive information regarding patient characteristics (including risk factors such as BMI, income, education, lifestyle, ethnicity,
ATSI status, CALD status) to permit risk adjustment (ACI, 2016; Sansoni, 2016; Smith & Street, 2013).

4b. PROMs data should include clinical diagnostic information, clinician-assessed outcome and other non-patient reported outcomes.

4c. PROMs data should be linked to comprehensive information regarding health service utilisation to facilitate economic evaluation and performance measurement (Sansoni, 2016).

4d. PROMs data should include patient self-report on OOP health care utilisation, return to work and participation in usual activities, and use of formal and informal home care to facilitate economic evaluation from a range of perspectives.
References


Hays et al. (2009). Development of physical and mental health summary scores from the patient-reported outcomes measurement information system (PROMIS) global items. *Qual Life Res*, 18: 873.


Revicki et al. (2009). Predicting EuroQol (EQ-5D) scores from the patient-reported outcomes measurement information system (PROMIS) global items and domain item banks in a United States sample. *Quality of Life Research*, 18: 783–91.


When: When to measure PROMs?

Background

The forgotten ‘W’?

- Rolfson and Malchau [1] suggested that:
  
  “the debate is not primarily why or if we should measure PROMs, but rather how, when and what to measure, and how to interpret the results.”

- While the ‘what’ and ‘how’ get most of the attention in the literature (e.g. [2, 3]), the ‘when’ tends to get overlooked: the Department’s PROMs consultation document for example does not discuss timing of PROMs measurement.

- This may be because clinical outcomes are generally evaluated after the patient outcome has stabilized, with the treatment deemed a success if the improvement in outcome is greater than some minimum important difference [1] – if this is the aim, the results are reasonably insensitive to timing – as long as enough time is allowed for the patient outcomes to stabilize post treatment.

- PROMs can and should be used for economic evaluation which has an explicit time dimension (the years in QALYs)

The importance of ‘when’

- Devlin and Appleby [4] noted:
  
  “inferring the benefit of treatment from just two observations of PROMs makes the timing of the second observation crucial. For example, collecting PROMs data six months after hip surgery might miss the time when patients first get back to their usual activities, as well as giving no real indication of the longer-term outcomes and durability.”

- Recent literature [5] showed bias in economic evaluations of total knee replacement due to timing of follow-up (Figure 1).
Economic evaluations using routinely collected PROMs will typically be pre-post evaluations which are susceptible to regression to the mean. This can also bias economic evaluations [6].

Not all follow-up points are of equal importance – in knee replacement, the patient trajectory after 12 months after surgery remains relatively flat [7]. This means that another follow-up at 4 years doesn’t add much value, where a follow-up at 3 months would. The SMART registry is now trialing this change [5].

Recommendations

1. Overarching principle for determining timing of PROMs measurements
   The overarching principle for the issue of timing should be that we maximize value from the PROMs in terms of benefits of outcome measurement data versus the costs and logistics of collecting the data.

   **Recommendation #1:** Evaluate value of measurement points (compare marginal benefit versus marginal cost of each follow-up point)

2. Minimum number of measurements
   On this basis we believe that there should be a minimum of two time points of data collection for each disease and/or treatment type. These are at pre- and post- hospital presentation time points to allow for assessment of the impact of treatment. Where a pre-treatment
measurement is not feasible (e.g. some acute events), the value of a post-treatment measurement should be evaluated according to recommendation #1. It is likely that such a measurement would still add value by allowing comparison between facilities in similar scenarios.

**Recommendation #2: Minimum 2 measurements, pre and post**

3. **Disease specific timing of measurement**
   For many diseases (e.g. hip/knee replacement surgery, fractures) there are known disease trajectories. Simple mathematical modelling should be used to determine the optimal timing for PROMs measurement based on existing data/research on disease trajectory. The pre-treatment measurement should be as close as possible to the treatment date (or admission date if there is no explicit treatment date). Once selected, the optimal timing should be adopted across all jurisdictions to avoid compromising comparisons and economic evaluations.

   There may be some value in choosing a common time point (12 months post intervention) for standardization/comparisons across disease/intervention, however this will depend on the objectives for collecting PROMs.

**Recommendation #3: Analyze disease trajectory to determine optimal measurement points – and as a result timing should be disease/treatment specific, but consistent across jurisdictions**

4. **Pilot period provides opportunity to test further measurements**
   In the pilot period there will be an opportunity to collect data at different time points for a given disease/intervention, to test what extra value is provided by such measurements. For example, a pre-treatment measurement at time of entry onto a waiting list might provide valuable information about the trajectory of patients towards treatment.

   Similarly, for non-elective treatments and/or chronic disease, the pilot period should be used to investigate the costs and benefits of regular measurements points.

**Recommendation #4: Use pilot to assess costs and benefits of further measurements (e.g. another pre-treatment measurement and/or another post-treatment measurement)**

5. **Measurement issues**
   Both the date and time of PROMs assessment must be collected to allow for adjustment for time of day outcomes measured, time from event (and to event in the case of elective
surgery). There are good time data already collected by hospitals related to the peri-operative period: this information needs to be linked to the PROMs.

**Recommendation #5:** Date and time of PROM measurement must be collected so can adjust time to/from event. Measurement should be made as close as possible to the designated measurement time point (e.g. 6 months post-treatment).

6. Further notes

**Chronic disease and non-elective treatments**

The timing of PROMs for elective surgery is more straight-forward than for chronic disease and non-elective treatments, which is perhaps why the UK have focused initially on elective treatments.

If the Department is considering including chronic disease, consideration needs to be given to value of obtaining PROMs data as per recommendation 1. For example there is probably no benefit in measuring PROMs in renal dialysis, but value is likely to be high in chronic obstructive pulmonary disease and chronic heart failure, i.e. those where chronic disease management interventions have been shown to be beneficial and where hospitals are introducing disease management interventions.

**Who is doing the data collection**

There is potential for bias and gaming due in data collection, and this can vary depending on if the collector is a hospital, a central agency or an externally commissioned network. This should be considered alongside the costs of each option. In the UK, hospitals collect pre-treatment and a central agency collects post-treatment. We think this is probably a reasonable compromise.

**Patient adaptation**

Patient adaptation (where patients change behavior in response to a condition) is an under-researched area that PROMs could help with. We recommend some thought into ‘future proofing’ for important research.
References


