Open access Original research

BMJ Open

Developing the People's Experience Survey (PES): a mixed-methods study updating a patient-reported experience measure (PREM) for use in any healthcare setting across Wales

Kathleen Withers [1,2] Robert Palmer,[1] Hawys Waddington,[1] Kathryn South,[3] Judith Lewis,[4] Richard Desir[5]

To cite: Withers K, Palmer R, Waddington H, *et al*. Developing the People's Experience Survey (PES): a mixed-methods study updating a patient-reported experience measure (PREM) for use in any healthcare setting across Wales. *BMJ Open* 2025;**15**:e100201. doi:10.1136/bmjopen-2025-100201

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (https://doi.org/10.1136/bmjopen-2025-100201).

Received 04 February 2025
Accepted 16 October 2025

[1]CEDAR, Cardiff and Vale University Health Board, Cardiff, UK
[2]Cardiff University, Cardiff, UK
[3]Value Transformation, NHS Wales Performance and Improvement, Cardiff, UK
[4]NHS Wales Shared Services Partnership, Cardiff, UK
[5]Office of the Chief Nursing Officer, Welsh Government, Cardiff, UK

Correspondence to
Dr Kathleen Withers;
kathleen.withers@wales.nhs.uk

ABSTRACT

Objectives To develop and validate a bilingual experience survey for use in any NHS healthcare setting, to support service improvement.

Design A prospective mixed-methods study.

Setting Any healthcare setting in NHS Wales including primary, secondary, urgent and planned care.

Participants An opportunistic sample of people with experience of using local healthcare services. Qualitative interviews and focus groups were held to develop a draft survey. These were followed by online data collection from a wide participant sample for statistical validation. The tool was translated and linguistically validated following recognised methods. Patient engagement leads were involved to ensure the tool met their needs.

Results We conducted and analysed five focus groups and four interviews, consisting of 33 people in total. 12 draft questions were developed related to key aspects of patient experience. A series of online surveys were conducted to test the draft questions, with 769 responses received. Data were analysed to assess completion rates, intra-rater reliability, internal consistency and convergent validity. One question had both sub-par intrarater reliability and poor convergent validity, and despite attempts to improve the wording, it failed to meet minimum requirements of validity and was subsequently removed. The final validated People's Experience Survey (PES) was subsequently translated into Welsh and validated with service users.

Conclusions The PES is a reliable and valid tool, suitable for use in any healthcare setting. The robust processes that have been undertaken ensure that the questions included are available bilingually to collect reliable, meaningful data to support service improvement work.

STRENGTHS AND LIMITATIONS OF THIS STUDY

- ⇒ Rigorous methods of validation ensure that the People's Experience Survey (PES) is capable of collecting meaningful, reliable data to support service improvements.
- ⇒ A wide range of service users was included, including groups that are traditionally hard to reach and those with specific healthcare needs, making it widely applicable.
- ⇒ Ongoing involvement of key stakeholders within healthcare ensured that the needs and requirements of national and local services could be met.
- ⇒ As a generic tool, the PES may be unable to pick up key issues for service users with specific healthcare requirements and may not be sensitive to condition-specific issues.

INTRODUCTION

Clinical outcomes such as mortality and readmissions have traditionally been used to assess the effectiveness of healthcare interventions. However, it is increasingly recognised that patient-reported data such as outcomes (PROs), experiences (PREs) and satisfaction can be used to assist services in meeting healthcare standards and are essential in measuring successful healthcare provision and effective person-centred care.[1 2]

While the term 'patient-reported experience' (PRE) is often used interchangeably with 'patient satisfaction', these are different constructs,[3] and patient satisfaction may reflect a person's expectations as much as the quality of care that they receive.[4] Patient-Reported Outcome Measures (PROMs) are validated tools, which measure a patient's perceived outcome of care by collecting PROs, which have widely accepted quality standards.[5 6] The validation process tests a tool's ability to accurately measure what it is designed to measure, ensuring the collection of robust, reliable data.[7] Although widely used, patient experience and satisfaction surveys, including Patient-Reported Experience Measures (PREMs), are often locally created for use in a particular department or service, and questions specific to that area are often included. Furthermore, often they are not validated. This means that there is a risk that the data they capture may not be generalisable across departments or hospitals or accurately reflect a patient's views, affecting their ability for use in informed decision-making processes. Increasingly, this is changing, with numerous PREMs now validated for use in a broad range of specific conditions and diseases.[8] The availability of a validated and generalisable PREM can facilitate high-quality data collection across services. This can support local and national service improvements and benchmarking by identifying areas with poor patient experience to facilitate improvement initiatives and areas with good experience to identify learning opportunities.

Some generic tools do exist, such as the Hospital Consumer Assessment of Healthcare Providers and Systems[9] used in the USA which was first implemented in 2006. However, this validated tool may not represent what is relevant to users of the varied, and free at point of use, healthcare settings of interest in NHS Wales. While the aims of the NHS Wales tool have some comparisons to the NHS England Friends and Family Test (FFT),[10] the FFT collects primarily qualitative feedback and it is not designed to support comparison across organisations.[11] Although there is no identified evidence of service user involvement in its development and validation, the tool has been widely adopted across NHS England since 2013 for service improvement initiatives and currently gathers approximately 2 million pieces of feedback a month.[10]

The recent international Patient-Reported Indicators Survey (PaRIS) initiative led by the Organisation for the Economic Co-operation and Development[12] was developed for use in the primary care setting. Designed for international use, it aimed to gather patient-reported information on the experiences and outcomes of primary care, particularly in relation to understanding the needs of people of 45 years or over, living with chronic conditions. This initiative included 19 countries (including Wales) that deployed the survey in 2023. While the target population of the PaRIS study is narrower than the NHS Wales programme, the widespread uptake illustrates the growing interest in gathering patient feedback to improve services.

Background

A core generalised patient experience survey was first launched in NHS Wales in 2013,[13] to collect consistent service user feedback to facilitate service improvement initiatives. Referred to as the 'National Core PREM', this survey was not a pure PREM, including categorical PREMs questions and questions requiring free-text responses. It was not intended to cover an exhaustive list of questions on all aspects that may impact on a service user's experience of care. Rather, it aimed to include questions on only the most relevant aspects related to a person's experience of care that would apply in any healthcare setting. The intention was to support consistent data collection across departments and services, while still allowing services to add bespoke questions relevant to their clinical area, if required. Following a service user engagement exercise, the National Core PREM was further updated in 2018, resulting in 11 recommended questions.[14]

The impact of COVID-19 led to a change in the way some healthcare services provide care across Wales, and the Welsh Government and health boards acknowledge that this has impacted on the way that people access and use services. This, in return, may affect their experience of care. It was therefore important to ensure that the national core PREM continues to reflect the experience aspects that are most important across the current range of healthcare settings, including planned and emergency care, primary care, maternity and mental health services. Consideration of service users including children and adults, and those with learning and physical disabilities and spiritual care needs was also key.

To confirm that the core set includes the most relevant and appropriately worded questions, a 'refresh' of the National Core PREM was requested.

As a result, the chief nursing officer at Welsh Government commissioned:

- Co-production and development of a generic validated bilingual set of national core questions applicable to all healthcare settings and commissioned services in NHS Wales.
- A review of the current Welsh Government Framework for Assuring Service User Experience in consultation with NHS and wider stakeholders, to align with current Welsh Government policies.

This manuscript details the first of these activities.

AIM

To develop a validated People's Experience Survey (PES) tool to act as a core set of PRE questions for use by people accessing any healthcare setting across NHS Wales.

Project outline

The project consisted of three distinct phases. Phase 1 aimed to develop a draft, English language, PES in a series of engagement events, while phase 2 aimed to test validity and reliability. A consultation period took place between these phases to gather feedback from the Safety & Learning Service User Feedback Network, a professional learning network managing patient safety and experience in NHS Wales. This was to align the draft PES with their needs and ensure it was fit for purpose. To ensure that patient-facing documents are available in both English and Welsh as per Welsh language requirements,[15] phase 3 aimed to robustly translate the final English language PES into Welsh, while maintaining the English version's validity and reliability properties.

METHODS

This project was led by a research team with experience in PROMs and PREMs development, with support from a steering group including patient experience leads from NHS Wales.

Phase 1: development of a draft tool

An existing version of a national PREM developed from work undertaken in 2013 and subsequently updated with service users in 2018[14] was used as a template to draft a new PES (see online supplemental material). This phase aimed to confirm the draft's content and face validity by engaging with service users from across Wales to:

- Identify redundant questions.
- Identify improvements in wording/phrasing.
- Identify additional areas or questions for inclusion. Where a new question was suggested, its wording was developed and tested with stakeholders.
- Finalise a draft survey for testing.

The 2018 PREMs development utilised a number of case studies to help stakeholders imagine what the experiences of other people might be in different care settings. These were reused but expanded to ensure that work aligned with the wider Welsh Government's Equality, Diversity and Inclusiveness agenda. To support this, areas such as maternity, mental health, children and young people, and spiritual care services were included. These were reviewed by the steering group, and the Lead Chaplain at a local health board as a spiritual care expert to ensure they were representative and inclusive.

Invitations to participate were sent out to 57 patient involvement and support groups, patient experience teams and community organisations from across Wales to invite them to participate in a series of focus groups and interviews. These were offered as both in-person and remote options in English or Welsh. An opportunistic sample of participants was included, based on the best availability of those participants who expressed an interest to take part in the focus groups, and all volunteers who were only available for interview. Prior to the sessions, attendees were provided with a participant information sheet and consent form. Sign language interpreters and support workers were included where required to facilitate inclusion.

A topic guide based on the existing core PREM was used to facilitate the discussions (see online supplemental material). Think-aloud techniques[16] were used during the feedback sessions, and each question from the 2018 PREM was discussed in turn, and its relevance and wording challenged, along with question order. Participants were asked to reconsider the questions while thinking of the case studies, and whether any additional considerations or questions might be needed to reflect the needs of service users. Subsequent focus groups were asked to consider the relevance and wording of any potential additions identified by previous focus groups.

Six focus groups (one in-person, five online) and four interviews (three online, one in-person) were held in total. Each focus group lasted approximately 2 hours, while interviews lasted approximately 90 min and were led by a researcher skilled and experienced in qualitative research methods (KW), with a second researcher acting as facilitator.

All focus groups and remote interviews were video-recorded and transcribed via Microsoft Teams. The in-person interview was audio-recorded and supported with field notes, then transcribed in an intelligent verbatim format. A thematic analysis was undertaken and emerging themes grouped together using inductive methods[17] to identify suggested questions and wording for use in the draft PES.

Phase 2: validity and reliability testing

The PES was further tested via a field test for:

- Completion rates: to test that the PES did not collect an abundance of missing data.
- Convergent validity (a form of construct validity): to confirm that the PES could measure the underlying construct of 'patient experience'.
- Intra-rater reliability: to ensure that the PES can measure something rather than collecting random data.

Internal consistency: to ensure that the PES questions can collect data consistent with each other and therefore measure aspects of the same underlying construct. Completion rates were assessed by calculating the proportion of responders who did not answer each question in the PES, while convergent validity was qualitatively estimated by comparing each response from a multiple-choice question in the PES with a corresponding free-text response, and judging their similarity in terms of semantics. While convergent validity is usually carried out by comparing the new tool with an existing 'gold-standard' one, as no relevant and validated alternative tools for measuring patient experience in NHS Wales were in existence, we propose that comparing to corresponding free-text responses was a reasonable alternative. This comparison for randomly selected responders was carried out by two researchers at Centre for Healthcare Evaluation Device Assessment and Research (CEDAR) followed by a reconciliation discussion. Intra-rater reliability was tested using the test–retest method, where participants are asked to complete the same questions about the same event on two occasions, with agreement indicating good reliability. A minimum threshold of ≥0.75 for the standard or weighted Cohen's kappa coefficient (or) κ_w[18] was considered acceptable. CIs[19] and p values (p) were also calculated, and pairwise deletion was used. We considered ordinal alpha (α)[20] values of ≥0.7 as a good measure of internal consistency, and confidence intervals calculated, and complete cases used.

It was not possible to test the PES' divergent validity, sensitivity nor responsiveness, as no reasonable assumptions could be made regarding expected patient experience and collecting data before and after a known service change was out of scope for this work.

Data were collected via online surveys which ran from 13 February 2024 to 25 March 2024. These included the draft PES questions identified in phase 1, plus additional demographic and free-text qualitative questions which mirrored those in the PES to provide context to the categorical responses and facilitate convergent validity testing. The 'test' survey was the first to be sent out, and to support the intrarater reliability testing, the same participants (with permission) were also asked to complete a second, 'retest' survey. Surveys were built on Microsoft Teams and sent to over 110 organisations and 300 individuals to complete. People were also asked to share the survey link with friends, colleagues and family members, leading to an opportunistic sample.

Phase 3: Welsh translation

Translating the final English PES into Welsh was facilitated by a researcher and a linguist experienced in translating PROMs, and by following internationally recognised principles of best practice.[21] This consisted of a team of five translators, with two independent forward translations from English to Welsh by qualified translators, and a reconciliation of these to yield a single Welsh version. This was followed by two back-translations into English by different translators, for comparison with the original source. The Welsh version was then validated through cognitive debriefing via interviews with seven Welsh-speaking patients and members of the public for feedback on its clarity, usability and reflection of the English version.

RESULTS

Phase 1 results

A convenience sample of 33 people joined in the discussions, with the youngest attendee aged between 18 and 25 and the oldest aged over 70; two-thirds (n=22) were female). All of the participants chose to engage using the English language. The attendees included service users, advocacy group representatives (ie, from support groups), and patient experience team representatives. Groups represented included Llais, Tenovus Cancer Care, Cwm Taf People First, Service User Involvement Group for Substance Support Services, British Deaf Association Wales and Cardiff Lupus Group. Self-reported health conditions included hearing impairment, sight impairment, mobility problems, issues with memory, stamina, dexterity and mental health.

General feedback

There were numerous discussions around the increasing volume of surveys currently collected both generally and by the NHS. In consideration of this, participants were keen that any request for patient-reported data should be kept to a minimum, with surveys kept short and concise, using simple language to improve readability. Accessibility was also discussed in relation to survey formats such as easy read and British Sign Language, as well as font type, size and colour. Allowing people to receive assistance when completing PREMs and similar surveys was generally seen as important, so that service users are able to get support from friends, family members or healthcare professionals, where required.

Good communication was acknowledged as heavily influencing the experience of any healthcare interaction, including the collection of patient-reported data such as experience and satisfaction. Participants provided strong feedback that in any data collection exercise, the introduction needs to be clear and specify why the data is being collected, and what it will be used for. A number of people noted that unless this information is explicit, they would not respond:

> I mean, I'm one of those people. I don't fill out a questionnaire unless you tell me what you're going to do with the data and how you're going to improve things. Service user, Focus Group 1.

Having a 'Thank You' at the end of a survey was also considered as important. Some participants suggested that being able to receive a copy of their responses would encourage them to take part in data collection exercises, particularly where more than one point of engagement is anticipated. People were keen to keep the number of questions included in the PES relatively low and thought the number of questions used in the 2018 PREM was appropriate. This acknowledged that careful consideration of questions was essential to reduce survey burden.

Themes

Existing questions

An introductory question of 'How recent was the experience you are thinking of?' was considered appropriate; however, the response options were updated to reflect the fact that feedback is now often collected in near real-time. Therefore, responses were changed to 'in the last week', 'between a week and a month ago', 'between 1 month and 6 months ago' and 'more than 6 months ago'.

The eight themes identified in previous work (10), and which corresponded to a question each in the initial survey (see table 3, questions 1–8) remained essential for inclusion. Participants thought that the wording of the original questions remained appropriate for four of these: being listened to; feeling cared for; having things explained in a way that was understood; and being involved in decisions about their care. However, the wording related to the other six themes was now considered inappropriate.

One of these questions was related to waiting times and read: 'From the time you needed to use this service, was the time you waited': with answer options of 'shorter than expected', 'about right', 'a bit too long' and 'much too long'. The theme of waiting times was thought to be very important as a concept related to patient experience. However, it was also thought to be particularly subjective due to the variability in people's expectations. There were also debates about how it is a more relevant issue for some services than others, such as emergency care or the Ambulance Service. The considerations discussed in relation to this theme included:

- The reference to a 'service' in the original question was too specific, particularly when trying to encompass different healthcare settings.
- People do not always realise that they need to use a service, but are referred on from elsewhere (eg, a general practitioner (GP) referring someone to secondary care).
- People may suspect they need to use a service but then wait for a period of time before they try to access it. This was thought to be a particular issue for people with chronic conditions who may try to self-care.

> Us people with chronic conditions, they're saying, "Well, I should call my GP, but I don't want to. I'll wait a week or so to see if it gets worse or it gets better". So THAT's when I realised that I needed to call the service, but I didn't. Service user, Focus Group 3.

Despite its challenges, most participants wanted the theme of 'waiting' included, and the wording was subsequently adapted to: 'Was the time you waited:' with answer options increased from four (two negative, one neutral and one positive) to five to prevent bias. This led to final answer options of 'much shorter than expected', 'a bit shorter than expected', 'about right', 'a bit longer than expected' and 'much longer than expected'.

Another question, 'If you asked for assistance, did you get it when you needed it?', had intended to relate to a broad range of help and support. However, feedback suggested that it was too vague to be useful, as there was too much room to interpret what 'assistance' might relate to, with some participants interpreting it as medical assistance and others seeing it as encompassing things like being given directions to a ward or being offered a wheelchair or an interpreter. 'Support' was offered as an alternative term, but this also had connotations of being related to something physical, like help with walking to the toilet. This meant that some participants would not have answered it in the intended way, and concerns from the patient experience teams around the ability to use this data to identify issues and improve services were seen as valid. Another issue with the wording was related to the consideration that people should not have to ask for assistance in all circumstances.

> [should] not necessarily have to ask for it, because if someone has it recorded in their records, that should be provided as a matter of course… If someone has a vision impairment then you would hope that people would realize that before someone came in, and make sure that whatever reasonable adjustments needed to be in place were in place for them to attend. Patient Advocate, Focus Group 4.

While there was some dissatisfaction with the wording, the general consensus was that the theme should be included with a question worded as 'If you needed help or support, did you get it when you needed it?'

'Were you able to speak in Welsh to staff if you needed to?' had been included in the 2018 tool partly to help measure health board adherence to the 2011 Welsh Language Measure.[15] This is legislation which places an obligation to provide services in Welsh and English on an equal basis. Language provision was seen in the engagement sessions as very important, but relevant on a much wider scale than Welsh and English, and also wider than the spoken word. There was very strong support to include a wider range of languages including British Sign Language, and to extend the question to incorporate general communication. Therefore, the question was broken into two parts, first asking about the preferred language of communication, with a list of options taken from the most commonly used languages in Wales as reported in the 2021 census.[22] These changes were discussed with health board equality teams to ensure they met legal requirements.

'Did you feel you understood what was happening in your care?' was considered to be very similar to 'Were things explained to you in a way that you could understand?', with patient experience teams noting that if people responded differently to questions, they would find acting on this challenging. In view of this, all groups agreed that this question should be removed.

An interval scale to rate overall experience, 'Using a scale of 0–10 where 0 is very bad and 10 is excellent, how would you rate your overall experience', was included in the initial PREM. This concept was popular, but the wording of the question itself was thought to be more cumbersome than necessary, and the number of response options too large.

> I do think you probably have too many options…. I think it needs to be reduced because if it's a seven, it might just as well be an eight. Patient Advocate, Focus Group 1

A reduced number of response options was also thought to be more practical from a user perspective for viewing the tool on different media such as mobile phone applications, where too many options may not all be visible at the same time. Five answer options were suggested as optimum and more aligned with the rest of the survey. There was also a preference for having labels instead of numbers to help people with choosing an appropriate response.

Free-text questions were thought to be important to give service users an opportunity to provide context to their answers, and two questions were felt to be appropriate. People liked the first of the two questions: 'Was there anything particularly good about your experience that you would like to tell us about?', however, there was a negative response to the second question: 'Was there anything that we could change to improve your experience?' Almost universally, participants thought this was inappropriately worded and that the NHS should 'be brave' and ask people simply and directly about anything that had been bad about their experience. In addition, a number of people noted that while they might have had a bad experience, they may not know how it could have been improved, and indeed, that making suggestions for improvement was not their responsibility.

Additional questions

Participants were very keen to keep the survey short to reduce survey fatigue. However, a number of new themes were suggested for inclusion, such as signposting to support services, accessibility, being supported in self-management and addressing concerns. While many of these were seen as important, they were not considered appropriate to include in a generic experience survey, as they were either not relevant in all care settings or were thought to be too similar in concept to existing questions. Reoccurring themes that were considered important across all focus groups and interviews were related to being treated with kindness, compassion, dignity and respect. An additional question of 'Were you treated with dignity and respect' was strongly supported by almost all participants as a means of measuring high-quality care, particularly in a landscape where NHS staff are under significant ongoing pressure.

> things like dignity and respect, kindness, being listened to, assistance. You know, those are your headliners……. I definitely think the dignity and respect one is really important. Patient Experience staff, Focus Group 3.

The questions were ordered based on feedback and appropriate headings agreed. This led to the production of a draft survey which was suitable for use in validity testing.

Phase 2 results

362 completed 'test' surveys were initially received, with responses from every county in Wales and representation from every demographic category (age group, gender and ethnicity). 'Retest' responses were received from 93 people, and of these, 65 could be linked back to their original 'test' response. The median time between test and retest responses was 7 days (range=5–21 days). Figure 1 illustrates the number of responders, linkage and the cohorts of responders used for testing. Participant demographics are available in the online supplemental material. Due to the recruitment methods of blanket invitations and snowballing (ie, asking people to share the invitation), it was not possible to identify how many people had received the invitation to participate, and therefore response rates could not be calculated.

Completion rates across all cohorts were high, with the highest proportion of missing data equal to 4% in cohort 3, question 6, as detailed in table 1, which details the number and percentage of each response option answered for every question by each cohort.

Convergent validity results were good, with most multiple-choice PES question responses matching their corresponding qualitative free-text response on the same theme (ie, a positive multiple-choice response matched a positive corresponding free text description). The exception was question 8, which did not show consistent results and was deemed unacceptable by the two qualitative researchers.

Inter-rater reliability was acceptable (κ or $\kappa_w \geq 0.75$) for all questions other than question 8, which had a k_w of 0.63 (see table 2, which shows the intrarater reliability of each question, measured via Cohen's kappa (κ) or weighted Cohen's kappa coefficient (κ_w) as appropriate). This question was subsequently discussed with the steering group and a small sample of service users, and it was agreed to attempt to adapt the wording of the question. It was split into two parts, with question 8a asking 'Did you need any help or support?', with answer options of 'Yes' or 'No'. If responders answered 'Yes', this was followed by question 8b: 'Did you get help and support when you needed it?", with answer options of 'always', 'usually', 'sometimes' and 'never'. This question was once again tested and retested with a new cohort of participants, and the intrarater reliability was subsequently recalculated using data from 81 responders, yielding a k, of 0.37 (0.16,0.59) for question 8a, which still did not meet the minimum acceptable threshold. Due to its poor convergent validity and intrarater reliability performance, it was decided to remove this question from the PES.

Internal consistency testing was subsequently carried out on all questions except question 8. This produced an ordinal alpha, α, value of 0.92, with a 95% CI of 0.78 to 0.98, well above the accepted reliability threshold of 0.7.

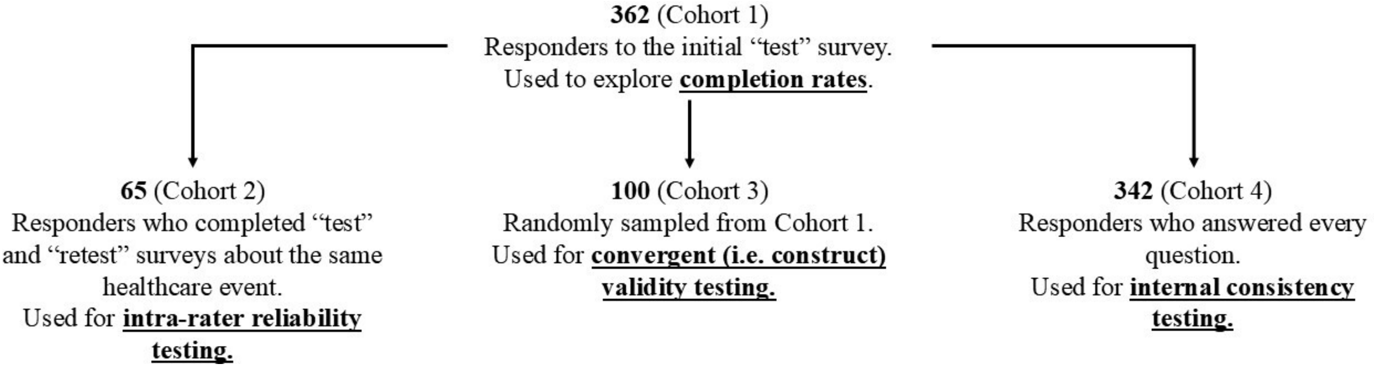

Figure 1 Respondent cohorts.

Table 1 Completion rates

Item	Answer options	Cohort 1 (n=362)	Cohort 2 (n=65) TEST	Cohort 2 (n=65) RETEST	Cohort 3 (n=100)	Cohort 4 (n=342)
Question 1: Was the time you waited:	Much shorter than expected	53 (14.6%)	16 (24.6%)	16 (24.6%)	16 (16.0%)	49 (14.3%)
	A bit shorter than expected	42 (11.6%)	10 (15.4%)	13 (20.0%)	12 (12.0%)	40 (11.7%)
	About right	117 (32.3%)	22 (33.8%)	19 (29.2%)	26 (26.0%)	112 (32.7%)
	A bit longer than expected	82 (22.7%)	9 (13.8%)	11 (16.9%)	26 (26.0%)	81 (23.7%)
	Much longer than expected	64 (17.7%)	8 (12.3%)	6 (9.2%)	18 (18.0%)	60 (17.5%)
	Unanswered	4 (1.1%)	0 (0%.0)	0 (0.0%)	2 (2.0%)	0 (0.0%)
Question 2: Did you feel well cared for?	Always	158 (43.6%)	41 (63.1%)	42 (64.6%)	52 (52.0%)	148 (43.3%)
	Usually	113 (31.2%)	16 (24.6%)	16 (24.6%)	19 (19.0%)	112 (32.7%)
	Sometimes	77 (21.3%)	8 (12.3%)	6 (9.2%)	25 (25.0%)	71 (20.8%)
	Never	11 (3.0%)	0 (0.0%)	0 (0.0%)	3 (3.0%)	11 (3.2%)
	Unanswered	3 (0.8%)	0 (0.0%)	1 (1.5%)	1 (1.0%)	0 (0.0%)
Question 3: Were you treated with dignity and respect?	Always	219 (60.5%)	52 (80.0%)	52 (80.0%)	59 (59.0%)	208 (60.8%)
	Usually	92 (25.4%)	10 (15.4%)	7 (10.8%)	21 (21.0%)	89 (26.0%)
	Sometimes	40 (11.0%)	3 (4.6%)	6 (9.2%)	17 (17.0%)	37 (10.8%)
	Never	8 (2.2%)	0 (0.0%)	0 (0.0%)	1 (1.0%)	8 (2.3%)
	Unanswered	3 (0.8%)	0 (0.0%)	0 (0.0%)	2 (2.0%)	0 (0.0%)
Question 4a: Were you able to communicate in your preferred language?	Always	327 (90.3%)	63 (96.9%)	62 (95.4%)	89 (89.0%)	312 (91.2%)
	Usually	13 (3.6%)	1 (1.5%)	2 (3.1%)	3 (3.0%)	12 (3.5%)
	Sometimes	12 (3.3%)	1 (1.5%)	1 (1.5%)	3 (3.0%)	12 (3.5%)
	Never	7 (1.9%)	0 (0.0%)	0 (0.0%)	2 (2.0%)	6 (1.8%)
	Unanswered	3 (0.8%)	0 (0.0%)	0 (0.0%)	3 (3.0%)	0 (0.0%)
Question 4b: Which language would you prefer to communicate in?	Welsh	22 (6.1%)	3 (4.6%)	2 (3.1%)	4 (4.0%)	21 (6.1%)
	English	332 (91.7%)	62 (95.4%)	63 (96.9%)	92 (92.0%)	315 (92.1%)
	Polish	1 (0.3%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (0.3%)
	Arabic	1 (0.3%)	0 (0.0%)	0 (0.0%)	1 (1.0%)	1 (0.3%)
	British Sign Language	3 (0.8%)	0 (0.0%)	0 (0.0%)	2 (2.0%)	2 (0.6%)
	Other	2 (0.6%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	2 (0.6%)
	Unanswered	1 (0.3%)	0 (0.0%)	0 (0.0%)	1 (1.0%)	0 (0.0%)
Question 5: Did you feel that you were listened to?	Always	183 (50.6%)	43 (66.2%)	43 (66.2%)	50 (50.0%)	176 (51.5%)
	Usually	111 (30.7%)	16 (24.6%)	13 (20.0%)	29 (29.0%)	106 (31.0%)
	Sometimes	54 (14.9%)	5 (7.7%)	9 (13.8%)	17 (17.0%)	53 (15.5%)
	Never	8 (2.2%)	0 (0.0%)	0 (0.0%)	2 (2.0%)	7 (2.0%)
	Unanswered	6 (1.7%)	1 (1.5%)	0 (0.0%)	2 (2.0%)	0 (0.0%)
Question 6: Were you involved as much as you wanted to be in decisions about your care?	Always	189 (52.2%)	45 (69.2%)	46 (70.8%)	52 (52.0%)	181 (52.9%)
	Usually	106 (29.3%)	13 (20.0%)	13 (20.0%)	29 (29.0%)	103 (30.1%)
	Sometimes	51 (14.1%)	5 (7.7%)	5 (7.7%)	13 (13.0%)	47 (13.7%)
	Never	12 (3.3%)	2 (3.1%)	1 (1.5%)	2 (2.0%)	11 (3.2%)
	Unanswered	4 (1.1%)	0 (0.0%)	0 (0.0%)	4 (4.0%)	0 (0.0%)
Question 7: Were things explained to you in a way that you could understand?	Always	218 (60.2%)	52 (80.0%)	51 (78.5%)	58 (58.0%)	210 (61.4%)
	Usually	99 (27.3%)	10 (15.4%)	8 (12.3%)	25 (25.0%)	95 (27.8%)
	Sometimes	33 (9.1%)	2 (3.1%)	6 (9.2%)	13 (13.0%)	30 (8.8%)
	Never	7 (1.9%)	1 (1.5%)	0 (0.0%)	1 (1.0%)	7 (2.0%)
	Unanswered	5 (1.4%)	0 (0.0%)	0 (0.0%)	3 (3.0%)	0 (0.0%)

Continued

Table 1 Continued

Item	Answer options	Cohort 1 (n=362)	Cohort 2 (n=65) TEST	Cohort 2 (n=65) RETEST	Cohort 3 (n=100)	Cohort 4 (n=342)
Question 8: If you needed help or support, did you get it when you needed it?	Always	132 (36.5%)	32 (49.2%)	32 (49.2%)	38 (38.0%)	124 (36.3%)
	Usually	90 (24.9%)	13 (20.0%)	9 (13.8%)	22 (22.0%)	87 (25.4%)
	Sometimes	67 (18.5%)	6 (9.2%)	7 (10.8%)	17 (17.0%)	63 (18.4%)
	Never	13 (3.6%)	1 (1.5%)	1 (1.5%)	4 (4.0%)	13 (3.8%)
	Not applicable	56 (15.5%)	13 (20.0%)	16 (24.6%)	16 (16.0%)	54 (15.8%)
	Unanswered	4 (1.1%)	0 (0.0%)	0 (0.0%)	3 (3.0%)	0 (0.0%)
Question 9: How would you rate your overall experience?	Very poor	10 (2.8%)	0 (0.0%)	1 (1.5%)	3 (3.0%)	10 (2.9%)
	Poor	31 (8.6%)	4 (6.2%)	4 (6.2%)	8 (8.0%)	30 (8.8%)
	Neither good nor poor	52 (14.4%)	2 (3.1%)	1 (1.5%)	20 (20.0%)	48 (14.0%)
	Good	140 (38.7%)	22 (33.8%)	17 (26.2%)	29 (29.0%)	136 (39.8%)
	Very good	128 (35.4%)	37 (56.9%)	42 (64.6%)	39 (39.0%)	118 (34.5%)
	Unanswered	1 (0.3%)	0 (0.0%)	0 (0.0%)	1 (1.0%)	0 (0.0%)

Welsh language

Towards the end of phase 2, feedback was received from health board Welsh language teams to request that questions 4a and 4b (‘Were you able to communicate in your preferred language?’ and ‘Which language would you prefer to communicate in?’) should be reversed. It was suggested that this would support equality monitoring. As this did not affect the statistical validity, it was feasible and changes were made, with face validity tested in a small number of service users and patient experience experts. This change was included in a final version of the PES.

Phase 3 results

Following internationally recognised guidelines,[21] multiple forward-translations, reconciliation and back-translations were carried out with no alterations. This was followed by cognitive debriefing with seven members of the Welsh-speaking public, where eight necessary alterations were identified to the reconciled version. These included using less formal sentence structures and simplifying long-winded and complicated sentences, using more common variants of spelling in Welsh, and changing the translation to match the English wording more closely.

DISCUSSION

Over 400 organisations and individuals from across Wales were contacted to take part in this project, with 33 people

Table 2 Intra-rater reliability

Item	Data type	N	Metric type	Metric value (95% CI)	P
Item 1: Was the time you waited:	Ordinal	65	k_w	0.78 (0.54 to 1.00)	≤0.005
Item 2: Did you feel well cared for?	Ordinal	64	k_w	0.83 (0.59 to 1.00)	≤0.005
Item 3: Were you treated with dignity and respect?	Ordinal	65	k_w	0.79 (0.56 to 1.00)	≤0.005
Item 4a: Were you able to communicate in your preferred language?	Ordinal	65	k_w	0.91 (0.66 to 1.00)	≤0.005
Item 4b: Which language would you prefer to communicate in?	Categorical	65	k	0.79 (0.56 to 1.00)	≤0.005
Item 5: Did you feel that you were listened to?	Ordinal	64	k_w	0.84 (0.60 to 1.00)	≤0.005
Item 6: Were you involved as much as you wanted to be in decisions about your care?	Ordinal	65	k_w	0.76 (0.52 to 0.99)	≤0.005
Item 7: Were things explained to you in a way that you could understand?	Ordinal	65	k_w	0.77 (0.53 to 1.00)	≤0.005
Item 8: If you needed help or support, did you get it when you needed it?	Ordinal (including a categorical ‘not applicable’ option)	65	k_w (with modified quadratic weights)	0.63 (0.44 to 0.83)	≤0.005
Item 9: How would you rate your overall experience?	Ordinal	65	k_w	0.87 (0.63 to 1.00)	≤0.005

taking part in phase 1 and over 769 responses received in phase 2. This, along with the methodological rigour, was a strength of our study, with responders from different groups with participants from all ages, ethnicities and counties across Wales represented. Involvement of a wide range of stakeholders, including service users, patient advocates and patient experience teams has ensured that the included questions represent the domains that are most important to service users in Wales.

After discussing a number of new and existing themes, thirteen questions were agreed, and these were tested in a wide population. Of these, 9 of the 10 multiple-choice questions were found to be valid and reliable, and these, along with the introductory question and two free-text questions, were included in a final version PES for Wales (see online supplemental material). The validated questions are available in table 3, which illustrates the question and answer options for both the 2018 Universal PREM set and the 2024 Patient Experience Survey.

The question which failed validation had been adapted from the 2018 PREM and had been subject to some discussion in the focus groups, as some participants questioned

Table 3 Comparison of 2018 and 2024 experience questions

2018 Universal PREMs questions			2024 patient experience survey		
Question		Answers	Question		Answers
1	Did you feel that you were listened to?	Always, usually, sometimes, never	1	Was the time you waited:	Much shorter than expected, a bit shorter than expected, about right, a bit longer than expected, much longer than expected
2	Did you feel well cared for?	Always, usually, sometimes, never	2	Did you feel well cared for?	Always, usually, sometimes, never
3	From the time you realised you needed to use this service, was the time you waited:	Shorter than expected, about right, a bit too long, much too long	3	Were you treated with dignity and respect?	Always, usually, sometimes, never
4	If you asked for assistance, did you get it when you needed it?	Always, usually, sometimes, never, not applicable	4a	Which language would you prefer to communicate in?	Welsh, English, Polish, Romanian, Panjabi, Urdu, Portuguese, Spanish, Arabic, Bengali, Gujarati, Italian, British Sign Language, other—please specify
5	Did you feel you understood what was happening in your care?	Always, usually, sometimes, never	4b	Were you able to communicate in your preferred language?	Always, usually, sometimes, never
6	Were things explained to you in a way that you could understand?	Always, usually, sometimes, never	5	Did you feel that you were listened to?	Always, usually, sometimes, never
7	Were you involved as much as you wanted to be in decisions about your care?	Always, usually, sometimes, never	6	Were you involved as much as you wanted to be in decisions about your care?	Always, usually, sometimes, never
8	Were you able to speak in Welsh to staff if you needed to?	Always, usually, sometimes, never, not applicable	7	Were things explained to you in a way that you could understand?	Always, usually, sometimes, never
9	Using a scale of 1–10, where 0 is very bad and 10 is excellent, how would you rate your overall experience?	1, 2, 3, 4, 5, 6, 7, 8, 9, 10	8	How would you rate your overall experience?	Very poor, poor, neither good nor poor, good, very good
10	Was there anything particularly good about your experience you would like to tell us about?		9	Was there anything particularly good about your experience you would like to tell us about?	
11	Was there anything that we could change to improve your experience?		10	Was there anything particularly bad about your experience you would like to tell us about?	

PREMs, Patient-Reported Experience Measures.

its ability to collect meaningful data due to its perceived lack of clarity. Indeed, the validation process proved this to be a relevant concern, due to the question's lack of both intrarater reliability and convergent validity, leading to its removal from the set. The identification and subsequent removal of this unreliable question illustrates the benefit of carrying out robust validation processes.

While validation within the field of PROMs is common, PREMs and Patient Experience tools are often developed locally with little or no validation undertaken. This was highlighted in a 2019 systematic review, which found that of 88 identified PREMs tools, only 20.5% had undergone appropriate reliability testing, and 3.4% had met accepted standards for criteria validity.[8]

While it is not possible to say that any tool is completely reliable or valid,[23] our study included numerous processes of evaluation which increase confidence that it is accurately measuring people's experience of care. However, despite the robust processes used in this study, several limitations remain. These include the potential biases in the sample population which consisted of a high number of people from patient groups and networks. While this was partially mitigated by inviting people from general community groups including substance support groups and third sector council members, the self-selecting nature of the project means that participants are likely to have been more engaged, and it is possible that we may not have captured views from people with a range of experiences. The recruitment method was chosen as it was considered to be the most effective way of engaging with a large volume of participants from a diverse network. However, this did prevent response rates from being calculated and also impacts on the ability to infer how many people will be likely to complete the PES while it is in use. Additionally, the nature and design of the study meant that it was not possible to assess the divergent validity, sensitivity or responsiveness of the tool.

While efforts were made to ensure that the PES is suitable for use in any healthcare setting, it is possible that it will lack sensitivity in its ability to identify challenges for some service users. Mental health service users, maternity service users and those with chronic conditions, for example, have very different needs, and access to care varies widely. There may also be challenges around ensuring that all service users are able to physically or remotely access the PES to provide feedback. While the aim of the PES is for clinical teams to use it as a basis with additional specific questions where appropriate, the core questions themselves may be less relevant to some patient groups.

While the PES is available bilingually, the population of Wales is diverse, and these languages may not represent all user needs. However, since the release of the PES, numerous additional translations have been completed, and the tool is available in all of the thirteen language options specified in the PES, including British Sign Language, and easy read options are also available. Different methods of dissemination/completion have also been implemented to improve accessibility. These include the use of SMS text messaging, paper versions, QR codes, online format and the use of kiosks/tablets, with SMS being the most effective form of method delivery across all age groups to date. This work supports inclusivity, with additional efforts underway to develop child-friendly versions of the tool and identify how hard-to-reach groups can be further engaged.

Despite the potential for challenges in implementing the PES across diverse settings, significant engagement with patient experience teams both during the development of the PES and since its launch has led to a strong uptake of the tool. It is now being used in all Health Boards and Trusts in Wales to support benchmarking initiatives. Currently, its use is primarily limited to secondary care, and further engagement work is required to improve its reach into primary and community services. Despite its wide dissemination, it remains to be confirmed whether the people who complete the tool will be representative of the population of Wales, and how effective it will be in facilitating service improvement initiatives.

Despite the ongoing challenges, over 70 000 responses have been collected in the first 5 months of implementation, and the survey is already emphasising the importance of analysing patient experience data through the lens of protected characteristics, particularly ethnicity. Health boards will use this data to inform targeted improvements. Thematic analysis of patient comments will enable health boards to identify disparities in areas like staff-patient relationships and care coordination. These efforts aim to embed equality monitoring and drive person-centred improvements across Welsh healthcare.

CONCLUSIONS

Gathering feedback from people about their experience of healthcare is an essential part of the quality improvement process; however, the usefulness of patient-reported data is limited by the quality of the data received. This project has developed and validated a set of experience questions which will facilitate the collection of meaningful, reliable patient-reported data. Use of the PES will be encouraged by Welsh Government. It will be used by NHS Wales as the primary means of collecting real-time experience data to drive service change to deliver equitable and quality services for our communities, with findings available to the population of Wales.

Acknowledgements We would like to thank everyone who supported this project by taking part in focus groups or interviews, as well as those who completed the experience surveys. We would also like to thank those individuals and groups who supported data collection by sharing the links to the surveys among their members and colleagues. This manuscript was prepared in accordance with the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) checklist.

Contributors Conceptualisation: KW, RP, RD, KS and JL; Methodology: KW and RP; Data collection: KW and RP; Formal analysis: KW and RP; PES Translation: HW and RP; Writing–original draft: KW and RP; Writing– review and editing: HW, RD, KS, JL. KW is the guarantor for this work.

Funding This work was supported by Welsh Government funding. RD is employed by Welsh Government; the funder did not influence the results/outcomes of the study despite author affiliations with the funder.

Competing interests None declared.

Patient and public involvement Patients and/or the public were involved in the design, or conduct, or reporting, or dissemination plans of this research. Refer to the Methods section for further details.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants. Prior to the start of the project, all relevant permissions were gained. The project was assessed by the Research and Clinical Audit teams and was identified as not being classed as a research or evaluation project. Therefore, no formal ethical or research approvals were required; however, senior organisational support was needed and attained from the Chief Executive of the lead organisation. In addition, advice was sought from the organisational Information Governance (IG) lead, and confirmation was received that no formal IG permissions were required to undertake the project. Recruitment took place between September 2023 and March 2024. Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request.

ORCID iD

Kathleen Withers https://orcid.org/0000-0001-9514-2025

REFERENCES

1. Gualandi R, Masella C, Piredda M, *et al*. What does the patient have to say? Valuing the patient experience to improve the patient journey. *BMC Health Serv Res* 2021;21:347.
2. NHS Wales Chief Medical Officer. Health and Care Quality Standards 2023. Welsh Health Circular (WHC/2023/013). 2023.
3. Bull C, Crilly J, Latimer S, *et al*. Establishing the content validity of a new emergency department patient-reported experience measure (ED PREM): a Delphi study. *BMC Emerg Med* 2022;22:65.
4. Anufriyeva V, Pavlova M, Stepurko T, *et al*. The validity and reliability of self-reported satisfaction with healthcare as a measure of quality: a systematic literature review. *Int J Qual Health Care* 2021;33.
5. Reeve BB, Wyrwich KW, Wu AW, *et al*. ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. *Qual Life Res* 2013;22:1889–905.
6. Mokkink LB, de Vet HCW, Prinsen CAC, *et al*. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. *Qual Life Res* 2018;27:1171–9.
7. Christensen KB, Comins JD, Krogsgaard MR, *et al*. Psychometric validation of PROM instruments. *Scandinavian Med Sci Sports* 2021;31:1225–38.
8. Bull C, Byrnes J, Hettiarachchi R, *et al*. A systematic review of the validity and reliability of patient-reported experience measures. *Health Serv Res* 2019;54:1023–35.
9. Services CfMM. Hospital Consumer Assessment of Healthcare Providers and Systems. CAHPS Hospital Survey. Baltimore, 2025. Available: https://www.hcahpsonline.org/en
10. NHS.UK. Friends and Family Test N.D, Available: https://www.england.nhs.uk/fft
11. NHS.UK. NHS England and NHS Improvement guidance: Using the Friends and Family Test to improve patient experience, 2019. Available: https://www.england.nhs.uk/publication/nhs-england-and-nhs-improvement-guidance-using-the-friends-and-family-test-to-improve-patient-experience
12. Bloemeke-Cammin J, Groene O, Ballester M, *et al*. International cross-cultural development and field testing of the primary care practice questionnaire for the PaRIS survey (PaRIS-PCPQ). *BMC Prim Care* 2024;25:168.
13. NHS Wales Shared Service Partnership. Patient Experience; Service user Experience. 2023. Available: https://nwssp.nhs.wales/a-wp/governance-e-manual/engaging-with-others/patient-experience/#:~:text=The%20Framework%20for%20Assuring%20Service,revised%20Health%20and%20Care%20Standards
14. Withers KL, Puntoni S, O'Connell S, *et al*. Standardising the collection of patient-reported experience measures to facilitate benchmarking and drive service improvement. *Patient Exp J* 2018;5:16–24.
15. National Assembly for Wales. Welsh Language (Wales) Measure, 2011. Available: https://www.legislation.gov.uk/mwa/2011/1/contents/enacted
16. Ericsson KA, Moxley JH. Thinking aloud protocols: Concurrent verbalizations of thinking during performance on tasks involving decision making. A handbook of process tracing methods for decision research. 2011;89–114.
17. Acquadro C, Conway K, Hareendran A, *et al*. Literature review of methods to translate health-related quality of life questionnaires for use in multinational clinical trials. *Value Health* 2008;11:509–21.
18. Fayers PM, Machin D. *Quality of life: the assessment, analysis and interpretation of patient-reported outcomes*. 2nd edn. John Wiley & Sons, 2013.
19. Fleiss JL, Levin B, Paik MC. *Statistical methods for rates and proportions*. 3rd edn. John Wiley & Sons, 2013.
20. Gadermann AM, Guhn M, Zumbo BD. Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide. *Pract Assessm Res Evaluat* 2019;17:3.
21. Wild D, Grove A, Martin M, *et al*. Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient-Reported Outcomes (PRO) Measures: Report of the ISPOR Task Force for Translation and Cultural Adaptation. *Value Health* 2005;8:94–104.
22. Office for National Statistics. Statistical Bulletin. Language, England and Wales: Census 2021. Main language, English language proficiency, and household language in England and Wales, Census 2021 data. 2021.
23. Frost MH, Reeve BB, Liepa AM, *et al*. What is sufficient evidence for the reliability and validity of patient-reported outcome measures? *Value Health* 2007;10 Suppl 2:S94–105.