Original Paper

Integrating Patient-Reported Outcome Measures Through the Health Outcomes Observatory (H2O) Digital Infrastructure: Multicenter Mixed Methods Feasibility Study

Preston Long[1], MA, MSc, PhD; Gemma Galan[2], MSc; Carolina E Watson[3], PhD; Yolima Cossio-Gil[4], MD, PhD; Margaret Renn Andrews[1], MPH; Valentin Ritschl[1,5], PhD; Valentina Schmolik[1], MSc; Marta Ferri[6], MSc; Ann-Kristin Porth[7], MD, PhD; Lisa Rebecca Otto[8,9], PhD; Alizé A Rogge[8,9], PhD; Linus Haberbosch[10], MD; Anouk Neureiter di Torrero[11], MSc; Laura Verbeij[11], MSc; Xavier Serra-Ruiz[12], MD; Vanesa Flores[13], MD; Marta Far Serrano[14], BSN, MSc; Rahim Lalji[15], PhD; Christian Primas[16], MD; Gottfried Novacek[16], MD; Tanja Stamm[1,5], MSc, MBA, PhD

[1]Institute for Outcomes Research, Center for Medical Data Science, Medical University of Vienna, Vienna, Austria

[2]Health Services Research Group, Vall d'Hebron Institut de Recerca, Vall d'Hebron Hospital Universitari, Barcelona, Catalonia, Spain

[3]Nursing Department, Faculty of Medicine, Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Catalonia, Spain

[4]Digital Strategy and Data Department, Hospital Sant Joan de Déu Barcelona, Barcelona, Catalonia, Spain

[5]Ludwig Boltzmann Institute for Arthritis and Rehabilitation, Vienna, Austria

[6]Digital Transformation Area, Vall d'Hebron Institut de Recerca, Barcelona, Catalonia, Spain

[7]Department of Internal Medicine III, Division of Endocrinology and Metabolism, Medical University of Vienna, Vienna, Austria

[8]Department of Psychosomatic Medicine, Charité – Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany

[9]Center for Patient-Centered Outcomes Research, Charité - Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany

[10]Department of Endocrinology, Diabetes and Metabolic Diseases, Charité – Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany

[11]Department of Surgery and Department of Quality and Patient Care, Erasmus MC, Rotterdam, South Holland, The Netherlands

[12]Gastroenterology Department, Crohn's and Colitis Attention Unit, Vall d'Hebron Hospital Universitari, Barcelona, Catalonia, Spain

[13]Endocrinology and Nutrition Department, Vall d'Hebron Hospital Universitari, Barcelona, Catalonia, Spain

[14]Medical Oncology Department, Vall d'Hebron Institute of Oncology, Vall d'Hebron Barcelona Hospital Campus, Barcelona, Catalonia, Spain

[15]Novartis Pharmaceuticals UK, London, United Kingdom

[16]Department of Internal Medicine III, Division of Gastroenterology and Hepatology, Medical University of Vienna, Vienna, Austria

Corresponding Author:
Gemma Galan, MSc
Health Services Research Group
Vall d'Hebron Institut de Recerca, Vall d'Hebron Hospital Universitari
Passeig de la Vall d'Hebron, 119-129
Barcelona, Catalonia
Spain
Phone: 34 934893000
Email: gemma.galan@sjd.es

Abstract

Background: The integration of patient-reported outcomes (PROs) into routine clinical practice remains limited despite their well-documented benefits in facilitating clinical communication and research. Systematic digital PRO collection can enable real-time symptom monitoring, timely clinical interventions, and greater patient empowerment. The Health Outcomes Observatory (H2O) initiative was developed to enable systematic digital patient-reported outcome measures (PROMs) collection and integration with clinical data to support value-based, patient-centered care across health care systems.

Objective: This study aimed to evaluate the feasibility, usability, acceptability, clinical utility, and implementation of the H2O digital infrastructure from patient and health care professional perspectives across multiple health care settings and medical conditions.

XSL•FO
RenderX

Methods: A noninterventional, multicenter, convergent mixed methods feasibility study was conducted in Spain, Austria, the Netherlands, and Germany across multiple medical conditions. Locally implemented smartphone, tablet, and web-based solutions captured and displayed PROs. Surveys and interviews assessed usability, acceptability, perceived clinical utility, and implementation conditions among patients and health care professionals.

Results: Of 241 participants assessed initially, 115 (47.7%) contributed follow-up data. Although 83% (44/53) of the respondents indicated willingness to recommend H2O, the mean System Usability Scale (SUS) score was 44.65/100 (SD 9.78), indicating poor usability. Among 50 participants with paired Patient Satisfaction Questionnaire Short Form assessments, satisfaction with technical quality declined (t_{49}=2.20; P=.03). Interviews identified perceived value alongside login difficulties, technical instability, weak integration with clinical systems, and inconsistent use of PRO data during consultations.

Conclusions: The H2O infrastructure could be deployed across heterogeneous settings, and participants perceived value in digital PRO collection; however, poor usability, substantial attrition, technical friction, and limited workflow integration qualify conclusions about feasibility and acceptability. More reliable technology, stronger interoperability, and evaluation in larger and more complete samples are needed before broader implementation.

(JMIR Form Res 2026;10:e94733) doi: 10.2196/94733

KEYWORDS

patient-reported outcomes; PROs; patient satisfaction; Health Outcomes Observatory; H2O; usability; digital health; patient-centered care; implementation; high-value care

Introduction

Background

Rising costs, fragmented care, aging populations, and chronic disease burden challenge health care sustainability [1]. Evaluating health outcomes is essential to assess system performance. Value-based health care (VBHC) focuses on health outcomes relative to the costs of achieving them [2]. Health care systems have begun implementing these principles [3], but patient-centered outcome data are not consistently collected in routine care. Performance assessment should include outcomes relevant to patients and the populations served [4].

Patient-reported outcomes (PROs) are one such crucial metric. PROs are direct reports from a patient about their health condition without influence by a clinician or anyone else [5]. One method of measuring these health aspects involves using validated patient-reported outcome measures (PROMs), which patients complete to assess the effects of symptoms, treatments, or both on their functional status, psychological well-being, social functioning [6], and overall health-related quality of life [7,8]. By capturing outcomes from the patient's perspective, PRO data can influence decision-making across multiple health care stakeholders, improve care pathways and health policies, inform health technology assessments, support medicine development and regulatory approval, and facilitate better communication between patients and health care professionals (HCPs) [5,8,9]. Evidence suggests that monitoring self-reported symptoms and ensuring timely responses from health care teams could lead to treatment adjustments and even improved patient survival [10].

Despite established use in clinical trials and medicine development [5,11], PROMs remain incompletely integrated into routine care. Paper collection and manual data entry introduce risks of transcription errors and delays, as documented in clinical data workflows [12-15]. PROM data can support both provider improvement and individual care, but these uses require appropriate collection and feedback arrangements [8,9]. Systematic and electronic patient-reported outcome measures (ePROMs) could enable real-time symptom monitoring, detect changes in disease progression, and facilitate timely interventions [16]. PROs should form a central component of the broader health outcome measurement framework [17-19], as their assessment is critical in promoting a patient-centered approach to health care. This approach empowers patients through self-management, engagement, and active participation in their care.

Clinical data, PROs, patient characteristics, and care-process data need to be considered together. Core outcome sets (COSs), developed through stakeholder consensus, standardize which outcomes should be collected for specific conditions [20,21]. Their practical value also depends on how those outcomes are collected and used.

The Health Outcomes Observatory (H2O) promotes VBHC through systematic collection and use of patient-relevant outcomes [22]. National observatories in Austria, Germany, the Netherlands, and Spain initially focused on diabetes, breast and lung cancer, and inflammatory bowel disease. This feasibility study evaluated patient and HCP experiences with the local technologies and workflows used to collect and review ePROMs in routine care.

Study Objectives

The primary objective was to evaluate the usability and perceived clinical utility of the national H2O ePROM infrastructure for routine care delivery from the perspectives of patients and HCPs. The secondary objectives were to identify the implementation factors most critical to successful ePROM integration into the diverse clinical practice settings and to understand current stakeholder acceptance, facilitators, and perceived barriers that could hinder implementation and scalability.

XSL•FO
RenderX

Methods

Design

This multicenter, prospective, pre-post study used a convergent mixed methods design with embedded qualitative components. Quantitative and qualitative data were collected in parallel with equal priority and integrated during interpretation. This feasibility design was selected as survey measures could quantify usability and satisfaction, while interviews could better explain the technical, workflow, and organizational conditions underlying quantitative results. One site in each of Austria, Germany, the Netherlands, and Spain participated, covering 1-4 indications per site. Local ethical approval was obtained at each site, and COS data collection was classified as noninterventional under local regulations. Additional protocol details are available elsewhere [23]. The mixed methods reporting of this study follows the Good Reporting of a Mixed Methods Study (GRAMMS) framework (refer to Multimedia Appendix 1 for the completed checklist).

H2O Digital Infrastructure and Workflow

H2O is a federated digital infrastructure. It combines disease-specific COS, locally selected patient-facing mobile or web apps, hospital dashboards or data-capture systems, and national governance arrangements. Site personnel enrolled and registered patients, provided training, and prompted completion of PROMs at baseline and scheduled follow-ups (electronic follow-up completion reminders were also provided). Patients completed questionnaires through the locally implemented app using a smartphone, tablet, or computer. Clinicians reviewed data through locally available dashboards or clinical tools; integration with electronic health records (EHRs) differed by site, and incomplete interoperability was itself an implementation limitation.

“H2O digital infrastructure” denotes the federated ecosystem, “local app” the patient-facing tool, and “study database” Clincase, which held feasibility evaluation data rather than supporting routine clinical review.

Collecting sites did not provide complete app names and versions, hosting and storage locations, EHR registration pathways, or PROM-to-EHR interfaces. These technical details could not be specified consistently across centers.

Operational Definitions

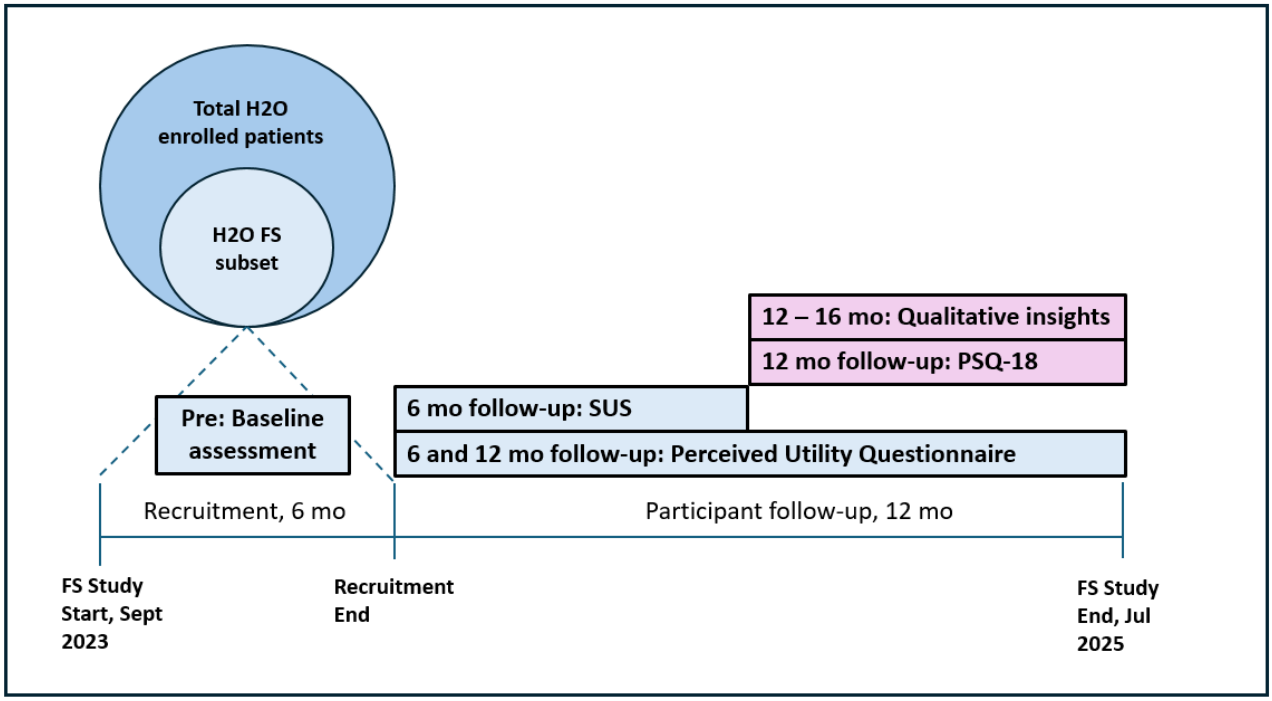
Feasibility was defined and operationalized as the ability to deploy and sustain digital ePROM collection under routine-care conditions across participating sites. Usability referred to perceived ease of use and interaction quality, assessed using the System Usability Scale (SUS) and process indicators. Acceptability referred to willingness to recommend and favorable or unfavorable experiences described in interviews. Implementation referred to the technical, workflow, staffing, and organizational conditions affecting deployment. Perceived clinical utility referred to whether patients and HCPs considered the collected PRO data useful for monitoring, communication, and clinical decision-making.

Each site used locally selected technology compatible with smartphones, tablets, or web browsers to capture and display PROs. Data architecture and access arrangements followed the relevant national H2O framework and local privacy and cybersecurity requirements. The degree of EHR integration differed by site, and some clinicians relied on separate dashboards or tools to review PROM data. Participation did not alter routine care; treatment decisions and disease monitoring remained at the treating physician’s discretion, and no additional study-specific clinic visits were required.

The study lasted 22 months, from first consent in September 2023 to final completion in July 2025 (Figure 1). Follow-up was planned for 12 months per participant, with study initiation and assessment timing varying according to site-level implementation of the COS and technology. Quantitative data assessed system usability, patient satisfaction with care, perceived utility, and process indicators. In parallel, interviews with patients and HCPs explored experiences, acceptability, and implementation context.

XSL•FO
RenderX

Figure 1. Planned assessment schedule for the Health Outcomes Observatory feasibility study. H2O: Health Outcomes Observatory; FS: Feasibility Study; mo: months; PSQ-18: Patient Satisfaction Questionnaire Short Form; SUS: System Usability Scale.



Implementation Disruptions and Protocol Deviations

Implementation and follow-up timing were not uniform across centers. One technology provider supporting an observatory became insolvent during data collection, delaying data retrieval. Other disruptions included prolonged technology implementation and troubleshooting, PROM licensing and translation barriers, and site-level modifications to planned procedures. Although the Patient Satisfaction Questionnaire Short Form (PSQ-18) follow-up was planned for 12 months, actual pre-post intervals ranged from 6 to 16 months. These deviations were treated as implementation constraints when interpreting feasibility and comparative results.

The collecting sites did not provide complete event-level records of vendor changes, software failures, protocol modifications, or dates of disruption and recovery. Consequently, the reported events cannot be assigned reliably to individual centers or linked quantitatively to outcomes. Site-level noncollection, including the absence of German PSQ-18 data, was distinguished from participant nonresponse when interpreting data availability.

Ethical Considerations

Participating observatories followed national legal and ethical governance arrangements. Consent materials and measures were adapted to local systems and languages while retaining common study procedures. All data transferred for analyses were received as pseudonymized and deidentified. Use of the data was limited to the scope of this study (ie, no secondary use). The master version of the informed consent, which was adapted for use by each site, is provided in Multimedia Appendix 2. Ethical approval was obtained from local institutional review boards of the Medical University of Vienna (EK1191/2023), Vall d'Hebron University Hospital (PR(AG)216/2023), Erasmus Medical Center (EC-2023-0735), and Charité University Hospital (EA4_034_22).

Baseline included the PSQ-18; SUS was scheduled at 6 months, the Perceived Utility Questionnaire at 6 and 12 months, PSQ-18 follow-up at 12 months, and interviews at 12-16 months. Actual paired PSQ-18 intervals ranged from 6 to 16 months; the schematic does not represent uniform achieved follow-up. Six patients per disease group and 3 HCPs specializing in each disease area were invited to interviews conducted 12-16 months after enrollment.

Study Population and Eligibility Criteria

Patients diagnosed with metastatic breast cancer (MBC), lung cancer (LC), Crohn disease (CD), ulcerative colitis (UC), or type 1 or type 2 diabetes mellitus (DM) who were receiving treatment at the Vienna General Hospital, Erasmus Medical Center, Vall d'Hebron University Hospital, and Charité University Hospital, along with their HCPs, were included in the study. Patients had to be at least 18 years old and able to read and understand the relevant local languages. Finally, all participants had to be willing and able to complete data collection using an electronic device. Each participant signed an informed consent form (or provided electronic confirmation) indicating understanding of agreement to participate in the feasibility study.

Study Sample

The study aimed to enroll a minimum of 200 patients, with 50 participants per site and at least 25 per disease group. Eligible participants had a diagnosis of type 1 or type 2 DM, CD, UC, MBC, or LC. No formal power calculation was performed for this exploratory study. Targets reflected historical clinic recruitment rates and the feasibility of obtaining descriptive and qualitative data.

XSL•FO
RenderX

Recruitment Procedure

HCPs and clinical staff at participating hospital centers recruited consecutive eligible patients during on-site clinical care, according to the usual clinical practice at each center. Informational materials supported the recruitment process and were reviewed with the patients. The HCP inviting the patient explained the purpose of the study and the role of PROMs and answered any questions before the patient decided whether to participate. Following this, interested patients provided written or electronic informed consent before participation. After consent, HCPs, with potential support from technology providers, guided patients on how to use the app and access the questionnaires. A purposive, maximum-variation sampling stratified by clinical condition was chosen. A subset of participants, stratified by age, gender, and education, was invited to participate in semistructured qualitative interviews. Patients could withdraw from the study at any time without explanation.

The purposive interview strategy described the intended sampling approach. The achieved sample was uneven across countries. Complete participant-level information on age, gender, education, and professional background was not provided by participants or collecting sites, preventing verification that the intended variation was achieved. The country-level demographic summaries available for the interviews are reported in the "Results" section.

Measures

The study collected disease-specific COS data, including PROs, clinical measures, and demographics. Additional feasibility study data included patient satisfaction with care and utility questionnaires, and qualitative interview responses. Data were collected from patients and HCPs through qualitative interviews and quantitative questionnaires (Multimedia Appendix 3).

SUS

The SUS assessed perceived usability. This widely used 10-item instrument covers ease of use, complexity, and confidence [24]. Items are rated on a 5-point scale from strongly disagree to strongly agree and converted using the standard scoring procedure to a total score from 0 to 100, with higher scores indicating better usability. SUS scores are scale scores, not percentages. For the descriptive interpretation used here, scores below 50 indicated poor usability and scores above 80 indicated very good usability [24]. These benchmarks describe perceived usability rather than clinical effectiveness or implementation readiness. Published evidence supports use of the SUS across technologies; however, internal consistency was not estimated for the present sample.

PSQ-18

The PSQ-18 assessed satisfaction with health care services. Its 18 items form 7 subscales covering general satisfaction, technical quality, interpersonal manner, communication, financial aspects, time spent with the doctor, and accessibility and convenience [25]. Items are rated on a 5-point scale and scored according to the published guidance, including reverse coding, so that higher scores indicate greater satisfaction. The PSQ-18 was planned at baseline and 12-month follow-up, although actual paired assessment intervals varied from 6 to 16 months. Internal consistency was not estimated for the present sample.

The PSQ-18 technical-quality subscale measures satisfaction with the technical quality of medical care; it does not directly measure the technical performance of software. Published measurement evidence does not establish validity in every clinical or language context. Complete item-level data needed to estimate internal consistency of SUS and PSQ-18 scores in this sample were not provided by the collecting sites; study-specific reliability and cross-language measurement equivalence could therefore not be evaluated.

Perceived Utility Questionnaire

The Perceived Utility Questionnaire (labeled by this name in Figure 1) was developed for the study to assess participants' experience of completing PROMs through the local H2O app. It collected gender, age group, and education and asked about willingness to recommend participation, ease of answering questions, time to completion, and ease of correcting responses using 0-10 numeric rating scales. It also recorded completion setting and device type. These items were used as indicators of practical usability, acceptability, and perceived utility; the questionnaire was not treated as a previously validated measurement instrument.

Age groups were collected in the Perceived Utility Questionnaire, whereas the demographic summaries report age in years as means and SDs.

Qualitative Interviews

Patient interview guides explored daily life, self-care, communication, use of PRO data during consultations, access barriers, and suggested improvements to training and app design. HCP guides covered clinical usefulness, patient-provider relationships, workflow integration, and implementation barriers. Both used open-ended questions on usability and perceived value.

Data Analysis

Quantitative Analysis

Descriptive statistics summarized the PSQ-18, SUS, Perceived Utility Questionnaire, and process indicators. SUS total scores, rather than subscales, were calculated. Bivariate correlation analyses assessed associations between participant characteristics and user-experience outcomes. One-way ANOVA compared device and completion-process groups. Paired 2-tailed *t* tests compared baseline and follow-up PSQ-18 subscale scores among participants with both assessments, with emphasis on technical quality, time spent with the doctor, accessibility and convenience, and communication. All inferential analyses were exploratory.

The reported *P* values are interpreted as exploratory, without a claim of confirmatory evidence across the multiple comparisons. The aggregate results do not permit independent verification of assumptions, model coding, site-adjusted effects, or all analysis-specific sample sizes. No causal inference was intended.

Study evaluation data were transferred to Clincase, a data-capture system hosted by the Medical University of Vienna;

XSL•FO
RenderX

some centers instead supplied Microsoft Excel files. Analyses were performed using IBM SPSS Statistics version 27 with a 2-sided α level of .05. Available-case analysis was used for each measure, and paired analyses included only participants with both assessments. No imputation or substitution was performed. A test for data missing completely at random was not conducted, and the analyses did not account for site-level clustering or potential nonrandom loss to follow-up; inferential results should therefore be interpreted cautiously.

Qualitative Analysis

The interview guides were developed by a multistakeholder project group containing HCPs, academic researchers, and industry partners from the H2O consortium. Qualitative data from semistructured interviews were analyzed thematically to explore user experiences and contextual factors influencing implementation.

Interviews were originally conducted in Spanish, Catalan, Dutch, or German, transcribed verbatim, and translated into English. Before API processing, the qualitative data were anonymized by deleting names and other identifying information, and interviewee answers were separated from interview questions and pseudonymization labels. Researchers summarized the interviewees' answers using GPT-4o-mini (OpenAI), accessed through the personal OpenAI API account of a researcher. Processing used controlled API access and the prepared text excluded identifying information; the research team reported that no data privacy issues were identified in this workflow.

Reflexive thematic analysis was used [26]. Two researchers (TS and MRA) independently reviewed and coded the full transcripts, compared interpretations, and resolved coding discrepancies through discussion; no discrepancies remained unresolved. Codes were iteratively refined into categories and themes addressing feasibility, usability, implementation, and perceived utility. Analytical memos and team discussions supported reflexivity and credibility. GPT-4o-mini was used to summarize interview responses; this use is disclosed separately from manuscript language editing.

Translation, removal of question context and labels, and AI-assisted summarization may reduce contextual nuance or obscure minority views. These processing steps should be considered when interpreting the themes; the use of AI does not itself establish the accuracy or completeness of the qualitative interpretation.

Mixed Methods Integration

To enhance the mixed methods approach, the research team (TS, PL, GG, and MRA) integrated quantitative and qualitative findings during interpretation using a weaving narrative strategy to identify convergence, complementarity, or divergence across data strands.

Integration concerned the interpretation of findings across data strands rather than estimation of an individual-level association between survey responses and interview themes. In particular, the predominantly Spanish interview sample could contextualize reported usability barriers but could not explain every country's quantitative results or the experiences of participants who did not provide follow-up data.

Results

Participant Flow and Data Completeness

Of 241 participants at the first assessment, 115 (47.7%) contributed any follow-up data, representing attrition of 126 (52.3%). Denominators differed across measures because of participant nonresponse and site-level noncollection. Completion setting, onboarding staff role, and device data were available for 120/241 (49.8%), 116/241 (48.1%), and 144/241 (59.8%) participants, respectively (Table 1). The SUS was completed by 205 participants, and 50 paired PSQ-18 records were available for the pre-post comparison. Missing data were analyzed using available cases without imputation. Relative to the 241 enrolled participants, completion-setting data were unavailable for 121 (50.2%), onboarding-role data for 125 (51.9%), device data for 97 (40.2%), and SUS data for 36 (14.9%). Fifty paired PSQ-18 records represented 20.7% of the enrolled cohort; this is an analysis-availability proportion, not a PSQ-18–specific dropout rate. Detailed enrollment, withdrawal, and noncompletion counts by site, disease, and scheduled time point were not provided by the collecting sites; so, a complete attrition matrix and reasons for individual noncompletion could not be reconstructed.

Across participants, the mean age was 46 (SD 14.4) years, and 130/241 (54%) were female (Table 2).

SUS data were available for 205 participants, and 195 unique PSQ-18 assessments were reported for the PSQ-18. Table 3 presents the supplied assessment counts by country. The interview column combines 26 patient and 8 HCP interviews.

Among records with process data, patients were most often recruited or onboarded by a physician or other HCP and completed PROMs on a smartphone at home or work (Table 1). Mean completion time was 231.34 (261.1) seconds (approximately 4 minutes; Table 4). Across 205 SUS respondents, the mean score was 44.65/100 (SD 9.78), indicating poor overall usability; site-level scores ranged from 26 to 86. No association with disease was reported. The German site did not collect PSQ-18 data.

XSL•FO
RenderX

Table 1. Process characteristics of patient-reported outcome measure (PROM) completion across diseases in the Health Outcomes Observatory feasibility study.

Variable and category	Values, n (%)
Process: setting (n=120)	
Waiting room	8 (6.7)
Doctor's office	8 (6.7)
Exam room	12 (10)
Hallway	2 (1.7)
Other	6 (5)
Home or work	84 (70)
Process: recruiting or onboarding staff role (n=116)	
Nurse	26 (22.4)
Clinician or physician	56 (48.3)
Admin	33 (28.4)
Other	1 (0.9)
Process: PROM device (n=144)	
Paper	1 (0.7)
Desktop	33 (22.9)
Laptop	18 (12.5)
Tablet	21 (14.6)
Smartphone	71 (49.3)

Table 2. Health Outcomes Observatory feasibility study participant characteristics by disease[a].

Disease	Values, n (%)	Age (years), mean (SD)	Male, n (%)
Total	241 (100)	46 (14.4)	106 (44)
Inflammatory bowel disease	124 (51)	42 (12.5)	69 (56)
Diabetes type I	42 (17)	37 (13.4)	10 (24)
Diabetes type II	29 (12)	64 (10.0)	27 (93)
Metastatic breast cancer	46 (19)	53 (8.3)	0 (0)

[a]Percentages in the table use the displayed disease-group totals as denominators and are rounded to whole numbers. Disease-specific male counts total 106. The narrative female count of 130 and the male count together account for 236 participants (missing n=5).

Table 3. Health Outcomes Observatory feasibility study assessment counts by country.

Country	SUS[a], n	PSQ-18[b], n	Interviews of patients, and HCPs combined[c], n
Austria	63	60	4
Germany	84	0	0
Netherlands	11	22	7
Spain	47	113	23

[a]SUS: System Usability Scale.

[b]PSQ-18: Patient Satisfaction Questionnaire Short Form.

[c]Interview counts combine patients and HCPs: 26 patients and 8 HCPs overall.

XSL•FO
RenderX

Table 4. Health Outcomes Observatory feasibility study usability testing descriptive statistics[a].

Measure	Minimum	Maximum	Mean (SD)
SUS[b] total score	0	100	44.65 (9.78)
Time to start (s)	0	120	2.35 (16.8)
Time to complete (s)	0	900	231.34 (261.1)
Ability to correct	1	4	1.78 (1.2)

[a]Time to start and time to complete are reported in seconds. The supplied "Ability to correct" summary uses a 1-4 code, whereas the questionnaire description refers to 0-10 ratings; its coding and interpretation could not be reconciled from the site information. It should not be interpreted as a count of errors or corrections.

[b]SUS: System Usability Scale.

Willingness to Recommend

Overall, 44/53 (83%) respondents indicated that they would recommend H2O. No association with disease, age, or sex was reported. Willingness to recommend differed across device groups in an omnibus comparison (n=49; $F_{3,4}$ =3.51, P=.02; η^2=0.190, 95% CI 0.002-0.343); this result alone does not establish a smartphone-specific difference. Of respondents willing to recommend H2O, 62% were smartphone users and 63% completed PROMs at home or work; these are distributions among recommenders, not within-group recommendation rates. Reported correlations with the device-use variable were r=0.80 for readability (P<.001), r=0.70 for response revision (P<.001), and r=−0.38 for initiation time (P=.005). Device coding and analysis-specific denominators were not supplied, so directional interpretations were not assigned. An association between SUS total score and willingness to recommend was reported (R^2=0.49, P=.009), but the model specification and sample size were unavailable for independent verification.

Change in Patient Satisfaction

The PSQ-18 comparison was limited to 50 participants with baseline and follow-up data. Follow-up was planned for 12 months, but actual intervals varied across sites from 6 to 16 months. Of the 7 PSQ-18 subscales, technical quality declined at follow-up (t_{49}=2.20, P=.03). No statistically significant changes were reported for general satisfaction, interpersonal manner, communication, financial aspects, time spent with the doctor, or accessibility and convenience, and no center-level difference was detected. Usability and onboarding staff role were not associated with patient satisfaction. Because the analysis used a small completer-only sample, these estimates are vulnerable to selection bias and should be interpreted cautiously.

Baseline and follow-up subscale means and SDs, paired mean changes, and their CIs were not supplied with the aggregate test results. The magnitude and clinical importance of the reported decline therefore cannot be determined from the P value alone. Nonsignificant comparisons should not be interpreted as evidence of equivalence or absence of an effect.

Qualitative Findings

Thirty-four interviews were conducted during the second half of the study: 26 with patients and 8 with HCPs. Of these, 23/34 (mean age 42, SD 13 years; 82% female) were conducted in Spain, 7/34 (mean age 66, SD 15 years; 57% female) in the Netherlands, 4/34 (mean age 46, SD 21 years; 50% female) in Austria, and none in Germany. The themes therefore primarily reflect Spanish implementation and should not be interpreted as representative of all 4 national infrastructures.

The available age and sex summaries were not accompanied by complete demographic denominators or separate patient and HCP profiles. Education, HCP professional roles, and a country-by-participant-type interview roster were not provided by participants or collecting sites. These omissions limit assessment of the achieved maximum-variation sample.

Patient Interviews

Seven patient themes were identified (Multimedia Appendix 4). Ease of navigation concerned participants' descriptions of simple navigation. Technical access and login problems concerned difficulty entering or using the local app. Reflection on physical and psychological health concerned the opportunity to consider symptoms and well-being when completing PROMs. Communication with the care team included frustration when responses were not discussed during appointments.

Perceived monitoring and research value concerned the potential usefulness of reporting health experiences and contributing data to research. Questionnaire design and personalization concerned restricted response options and requests for more personalized and engaging questionnaires. Unmet mental health support concerned participants' requests for greater attention to psychological needs. These findings describe reported experiences and preferences, rather than demonstrated improvements in clinical outcomes.

HCP Interviews

Five HCP themes were identified (Multimedia Appendix 4). Clinical value of PROs concerned the additional view of patients' health experiences offered by PROM collection. Identification of psychosocial and adherence concerns concerned issues that might otherwise remain unrecognized in routine consultations.

Workflow and EHR integration concerned poor synchronization with clinical tools, limited review of responses, and missed opportunities to discuss findings. Technical and digital-inclusion barriers concerned usability difficulties, particularly for older or less digitally experienced patients. Staff training and organizational support concerned the need for staff preparation and explicit routines for reviewing and acting on PROM data.

XSL•FO
RenderX

HCPs supported PROM integration in principle while identifying practical conditions needed for routine use.

Discussion

Principal Findings

The findings from this H2O feasibility study present a mixed picture. There is positive evidence for the perceived utility and acceptability of the H2O system from both patients and HCPs, while technical usability of the system was unsatisfactory. Willingness to recommend was retained as one global indicator of acceptability [27,28], but it should not be interpreted as a proxy for usability. Although 44/53 (83%) respondents indicated willingness to recommend H2O, the mean SUS score of 44.65 (SD 9.78) indicated poor usability; technical-quality satisfaction declined among the 50 paired PSQ-18 respondents. While the collection of PROMs via smartphone was predictive of willingness to recommend, further testing is required to accurately interpret this relationship. These results, together with workflow disruption and uneven site participation, limit conclusions about sustained use and require cautious interpretation beyond the participating settings and completer-only sample.

Patients valued reflection on health and research participation, consistent with earlier PROM studies [29-31]. Login difficulties, repetitive questions, and limited response options echoed reported barriers involving questionnaire fatigue and digital literacy [29,32]. Secure authentication remains necessary, but access should be usable. These findings primarily reflect that the Spanish interview sample of HCPs recognizes the value of collecting PROMs to better understand patients' health experiences and improve care quality. However, challenges remain, including difficulty reviewing responses in a timely manner, limited usability, and poor integration with existing clinical tools. Integrating PROM data into EHR was emphasized as crucial for facilitating consultations and informed discussions. Patient engagement, especially among older or less tech-savvy individuals, highlights the need for staff training to effectively support PROM use. These findings align with previous qualitative studies and systematic reviews reporting similar barriers and the importance of training and system integration for successful PROM implementation [33-35].

Integration of Quantitative and Qualitative Findings

The quantitative and qualitative strands converged in showing perceived value alongside substantial technical friction. Interviews helped explain how willingness to recommend could coexist with low SUS scores: participants valued remote reporting, reflection, communication, and the purpose of PROM collection, while describing login problems, unstable technology, limited clinical follow-up, and weak integration with routine tools. HCP accounts similarly contextualized site-level variability through workflow constraints, staffing, training needs, and limited interoperability. The strands therefore support conditional, rather than unqualified, feasibility; that is, perceived value depended on reliable technology and meaningful use of PROM data in care.

The mean SUS score of 44.65 (SD 9.78) was in the poor range. Its association with willingness to recommend does not offset the low absolute usability rating; its direction requires confirmation of recommendation coding and model specification. Variation across sites may reflect differences in local apps and implementation, but causal attribution is not possible without adjusted analyses accounting for site and missingness. Usability optimization remains a central requirement for future PROM systems.

Among the 50 participants with paired PSQ-18 data, satisfaction with the technical quality of medical care declined. This domain does not directly assess software reliability, so the finding cannot be equated with deterioration in platform performance. Interviews identified technology and workflow problems, but these accounts do not establish the cause of the PSQ-18 change. The small completer-only sample, variable 6- to 16-month intervals, multiple comparisons, and absence of adjustment for site or nonrandom missingness limit interpretation. Nonsignificant changes in other domains do not establish unchanged satisfaction.

Taken together, the findings identify usability and technical reliability as priorities for future implementation. Future studies should evaluate local apps across devices and settings using larger, more complete datasets and clearly documented assessment schedules. Whether such changes improve patient engagement or satisfaction with care requires further evaluation.

Strengths and Limitations

The multinational, cross-disease design enabled evaluation across heterogeneous clinical and technical contexts and helped identify recurring implementation challenges. However, transferability beyond the participating settings is limited by substantial attrition, uneven site and disease representation, incomplete assessment coverage, and the geographic imbalance of the qualitative sample.

Provider insolvency, implementation delays, and licensing and translation barriers produced uneven technology exposure and follow-up conditions.

Recruitment and data collection depended on HCP engagement and local staffing, technical expertise, and turnover. These constraints contributed to inconsistent sample sizes, missing data, and variable measurement timing. Attrition was 52.3%; only 50 paired PSQ-18 records were analyzed, the German site did not collect PSQ-18 data, and no missing-completely-at-random test, imputation, sensitivity analysis, or adjustment for site clustering was performed. The qualitative sample was also imbalanced: 23/34 interviews came from Spain, 7 from the Netherlands, 4 from Austria, and none from Germany. The themes therefore primarily reflect Spanish workflows and should not be generalized to all 4 national infrastructures.

The national observatories were established independently of this study, and detailed site-specific architecture, vendor-change logs, software-failure records, and protocol-deviation timelines were not provided by the collecting sites. The effect of individual disruptions could therefore not be quantified, which limits reproducibility. Study-specific internal consistency

XSL•FO
RenderX

estimates, completer-noncompleter comparisons, and PSQ-18 effect estimates and CIs were also unavailable; the device ANOVA includes η^2 and its 95% CI.

Mixed methods integration was limited by unequal coverage of countries, time points, and participants across the 2 strands. Interview accounts cannot establish why noncompleters discontinued participation or explain quantitative results from Germany, where no interviews were collected. Translation and AI-assisted summarization may also have altered nuance, particularly after removal of interview questions and labels. Apparent convergence should therefore be interpreted as contextual support rather than independent validation of survey estimates.

Conclusion

The H2O digital infrastructure could be deployed across heterogeneous settings, and patients and HCPs described value in systematic digital PROM collection. However, willingness to recommend should be interpreted alongside poor overall usability, declining technical-quality satisfaction in a small paired sample, substantial attrition, uneven site participation, and limited interoperability. The findings support conditional feasibility rather than broad implementation readiness. Reliable technology, consistent follow-up, integration of PROM results into clinical workflows, and adequately powered analyses that address clustering and missingness are needed before effectiveness or generalizability can be established. The successful routine collection of PROMs, which depends substantially on the ease of use and usability of the underlying technology, is essential for effectively incorporating the patient perspective into clinical care. Furthermore, the collection of PROMs across diseases and national borders offers substantial potential for research, including comparative evaluations of treatment outcomes. However, realizing this potential requires robust, well-implemented, and adaptable systems that are practical for routine clinical use.

Funding

The authors declare that financial support was received for the research and/or publication of this article. This work was supported by the Innovative Medicines Initiative (IMI) 2 Joint Undertaking under grant agreement number 945345-2. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme, the European Federation of Pharmaceutical Industries and Associations (EFPIA), Trial Nation, and the Juvenile Diabetes Research Foundation (JDRF). The public grant funding is matched with in-kind contributions of EFPIA partners. IMI is a partnership between the European Union and the European pharmaceutical industry, represented by the EFPIA. It is working to improve health by speeding up the next generation of medicines, particularly in areas where there is an unmet medical or social need. It works by facilitating collaboration between the key players involved in health research, including universities, research centers, pharmaceutical and other industries, small and medium-sized enterprises, patient organizations, and medicines regulators. IMI is the world's most extensive public-private partnership in the life sciences. All authors of this paper are members of the H2O. The funder had no role in the conception of this work. This manuscript reflects only the authors' view. The European Union and IMI are not responsible for any use that may be made of the information it contains.

Data Availability

The anonymized datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Generative AI Use

GPT-4o-mini (OpenAI) was used through a researcher's OpenAI API account to summarize translated, anonymized interview responses, as described in the "Qualitative Analysis" section. ChatGPT (GPT-5; OpenAI) was used during manuscript preparation and revision to reduce word count, improve wording, and formatting. AI-assisted qualitative summarization was distinct from manuscript editing. The authors retain responsibility for the study's analysis, interpretation, accuracy, and final content.

Authors' Contributions

Conceptualization: PL (lead), TS (supporting)
Data curation: PL (lead), GG, MRA, AKP, VS, VR (supporting)
Formal analysis: PL (lead), GG, TS, MRA, RL, LV, VS, MF (supporting)
Funding acquisition: TS
Investigation: YC-G, LH, XS-R, VPFE, MFS, CP, GN, AKP
Methodology: PL (lead), LRO, AAR, ANdT, VR, LV (supporting)
Project administration: PL (lead), GG (supporting)
Resources: TS, GN, CP, MFS, VPFE, YC-G, XS-R
Software: PL, LRO, AAR, LV, VR
Supervision: TS
Validation: CEW, MRA, ANdT
Visualization: RL, VS

XSL•FO
RenderX

Writing – original draft: PL (lead), RL, GG (supporting)
Writing – review & editing: TS, VS, MRA, GN

Conflicts of Interest

TS reports grant support from Roche and personal fees from AbbVie, Roche, Sanofi, Takeda, Novartis, and AOP Health, outside the submitted work. CP received lecture fees/consultant fees/advisory board member fees from AbbVie, Alfasigma, MSD, Takeda, Janssen-Cilag, Eli Lilly, Pfizer, Falk, Roche, Galapagos, Ferring, AstroPharma, and Amgen. GN received lecture fees/consultant fees/advisory board member fees from AbbVie, Merck Sharp & Dohme, Takeda, Janssen-Cilag, MEDahead, Pfizer, Gilead, Galapagos/Alfasigma, Ferring, Vifor, AstroPharma, Falk, Bristol Myers Squibb, Arena Pharmaceuticals, outside the submitted work. RL is currently an employee of Novartis. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Multimedia Appendix 1

GRAMMS (Good Reporting of A Mixed Methods Study) checklist used to guide reporting of the study.
[DOCX File , 30 KB-Multimedia Appendix 1]

Multimedia Appendix 2

Health Outcomes Observatory feasibility study patient consent form.
[DOCX File , 65 KB-Multimedia Appendix 2]

Multimedia Appendix 3

Utility questionnaire and interview guides for patients and health care providers.
[DOCX File , 18 KB-Multimedia Appendix 3]

Multimedia Appendix 4

Patient and health care professional interview themes.
[DOCX File , 2458 KB-Multimedia Appendix 4]

References

1. Porter ME, Teisberg EO. Redefining Health Care: Creating Value-Based Competition on Results. Boston. Harvard Business School Press; 2006.
2. Porter ME. What is value in health care? N Engl J Med. 2010;363(26):2477-2481. [doi: 10.1056/NEJMp1011024] [Medline: 21142528]
3. Withers K, Palmer R, Lewis S, Carolan-Rees G. First steps in PROMs and PREMs collection in Wales as part of the prudent and value-based healthcare agenda. Qual Life Res. 2021;30(11):3157-3170. [FREE Full text] [doi: 10.1007/s11136-020-02711-2] [Medline: 33249539]
4. de Bienassis K, Kristensen S, Hewlett E, Roe D, Mainz J, Klazinga N. Measuring patient voice matters: setting the scene for patient-reported indicators. Int J Qual Health Care. 2021;34(Suppl 1):ii3–ii6. [doi: 10.1093/intqhc/mzab002] [Medline: 33575802]
5. Guidance for Industry. Patient-reported outcome measures: use in medical product development to support labeling claims. US Food and Drug Administration. 2009. URL: https://www.fda.gov/media/77832/download [accessed 2026-10-13]
6. McGrail K, Bryan S, Davis J. Let's all go to the PROM: the case for routine patient-reported outcome measurement in Canadian healthcare. Healthc Pap. 2011;11(4):8-18. [doi: 10.12927/hcpap.2012.22697] [Medline: 22543287]
7. Coulter A. Measuring what matters to patients. BMJ. 2017;356:j816. [doi: 10.1136/bmj.j816] [Medline: 28219884]
8. Black N. Patient reported outcome measures could help transform healthcare. BMJ. 2013;346:f167. [FREE Full text] [doi: 10.1136/bmj.f167] [Medline: 23358487]
9. Devlin NJ, Appleby J. Getting the Most Out of PROMs: Putting Health Outcomes at the Heart of NHS Decision-Making. London. The King's Fund; 2010.
10. Basch E, Deal AM, Kris MG, Scher HI, Hudis CA, Sabbatini P, et al. Symptom monitoring with patient-reported outcomes during routine cancer treatment: a randomized controlled trial. J Clin Oncol. 2016;34(6):557-565. [doi: 10.1200/JCO.2015.63.0830] [Medline: 26644527]
11. Calvert M, Kyte D, Mercieca-Bebber R, Slade A, Chan A, King MT, the SPIRIT-PRO Group, et al. Guidelines for inclusion of patient-reported outcomes in clinical trial protocols: the SPIRIT-PRO extension. JAMA. 2018;319(5):483-494. [FREE Full text] [doi: 10.1001/jama.2017.21903] [Medline: 29411037]

XSL•FO
RenderX

12. Bauer J, John E, Wood C, Plass D, Richardson D. Data entry automation improves cost, quality, performance, and job satisfaction in a hospital nursing unit. J Nurs Adm. 2020;50(1):34-39. [doi: 10.1097/NNA.0000000000000836] [Medline: 31804410]
13. Vawdrey DK, Gardner RM, Evans RS, Orme JF, Clemmer TP, Greenway L, et al. Assessing data quality in manual entry of ventilator settings. J Am Med Inform Assoc. 2007;14(3):295-303. [FREE Full text] [doi: 10.1197/jamia.M2219] [Medline: 17329731]
14. Vecellio E, Maley MW, Toouli G, Georgiou A, Westbrook J. Data quality associated with handwritten laboratory test requests: classification and frequency of data-entry errors for outpatient serology tests. Health Inf Manag J. 2015;44(3):7-12. [doi: 10.1177/183335831504400302]
15. Meccariello M, Perkins D, Quigley L, Rock A, Jiejing Q. Vital time savings: evaluating the use of an automated vital signs documentation system on a medical/surgical unit. J Healthc Inf Manag. 2010;24(4):46-51. [FREE Full text]
16. Kotronoulas G, Kearney N, Maguire R, Harrow A, Di Domenico D, Croy S, et al. What is the value of the routine use of patient-reported outcome measures toward improvement of patient outcomes, processes of care, and health service outcomes in cancer care? A systematic review of controlled trials. J Clin Oncol. 2014;32(14):1480-1501. [FREE Full text] [doi: 10.1200/JCO.2013.53.5948] [Medline: 24711559]
17. Browne J, Cano S, Smith S. Using patient-reported outcome measures to improve health care: time for a new approach. Med Care. 2017;55(10):901-904. [doi: 10.1097/MLR.0000000000000792] [Medline: 28857961]
18. Boyce MB, Browne JP, Greenhalgh J. The experiences of professionals with using information from patient-reported outcome measures to improve the quality of healthcare: a systematic review of qualitative research. BMJ Qual Saf. 2014;23(6):508-518. [FREE Full text] [doi: 10.1136/bmjqs-2013-002524] [Medline: 24505110]
19. Boyes A, Newell S, Girgis A, McElduff P, Sanson-Fisher R. Does routine assessment and real-time feedback improve cancer patients' psychosocial well-being? Eur J Cancer Care (Engl). 2006;15(2):163-171. [doi: 10.1111/j.1365-2354.2005.00633.x] [Medline: 16643264]
20. Williamson PR, Altman DG, Bagley H, Barnes KL, Blazeby JM, Brookes ST, et al. The COMET handbook: version 1.0. Trials. 2017;18(Suppl 3):280. [FREE Full text] [doi: 10.1186/s13063-017-1978-4] [Medline: 28681707]
21. Kalf RRJ, Vreman RA, Delnoij DM, Bouvy ML, Goettsch WG. Bridging the gap: can international consortium of health outcomes measurement standard sets align outcomes accepted for regulatory and health technology assessment decision-making of oncology medicines. Pharmacol Res Perspect. 2021;9(2):e00742. [FREE Full text] [doi: 10.1002/prp2.742] [Medline: 33749172]
22. Stamm T, Bott N, Thwaites R, Mosor E, Andrews M, Borgdorff J, et al. Building a value-based care infrastructure in Europe: the health outcomes observatory. NEJM Catal Innov Care Deliv. 2021;2(3). [FREE Full text]
23. A prospective, multi-country, multi-centre, multi-condition feasibility study evaluating the establishment of the national health outcomes observatory (H2O) ecosystem. Health Outcomes Observatory. 2025. URL: https://zenodo.org/records/12624969 [accessed 2026-10-29]
24. Lewis JR. The System Usability Scale: past, present, and future. Int J Hum-Comput Interact. 2018;34(7):577-590. [FREE Full text] [doi: 10.1080/10447318.2018.1455307]
25. Thayaparan AJ, Mahdi E. The Patient Satisfaction Questionnaire Short Form (PSQ-18) as an adaptable, reliable, and validated tool for use in various settings. Med Educ Online. 2013;18:21747. [FREE Full text] [doi: 10.3402/meo.v18i0.21747] [Medline: 23883565]
26. Braun V, Clarke V. Thematic analysis. In: A Practical Guide. London. SAGE Publications Ltd; 2022.
27. Otani K, Waterman B, Faulkner KM, Boslaugh S, Dunagan WC. How patient reactions to hospital care attributes affect the evaluation of overall quality of care, willingness to recommend, and willingness to return. J Healthc Manag. 2010;55(1):25-37. [Medline: 20210071]
28. Klinkenberg W, Boslaugh S, Waterman B, Otani K, Inguanzo J, Gnida J, et al. Inpatients' willingness to recommend: a multilevel analysis. Health Care Manage Rev. 2011;36(4):349-358. [doi: 10.1097/HMR.0b013e3182104e4a] [Medline: 21685795]
29. Campbell R, Ju A, King MT, Rutherford C. Perceived benefits and limitations of using patient-reported outcome measures in clinical practice with individual patients: a systematic review of qualitative studies. Qual Life Res. 2022;31(6):1597-1620. [doi: 10.1007/s11136-021-03003-z] [Medline: 34580822]
30. Carfora L, Foley CM, Hagi-Diakou P, Lesty PJ, Sandstrom ML, Ramsey I, et al. Patients' experiences and perspectives of patient-reported outcome measures in clinical care: a systematic review and qualitative meta-synthesis. PLoS One. 2022;17(4):e0267030. [FREE Full text] [doi: 10.1371/journal.pone.0267030] [Medline: 35446885]
31. Consolo L, Basile I, Colombo S, Rusconi D, Pasquot L, Campa T, et al. Exploring patient perspectives on electronic patient-reported outcome measures in home-based cancer palliative care: a qualitative study. Digit Health. 2024;10:20552076241249962. [FREE Full text] [doi: 10.1177/20552076241249962] [Medline: 38665884]
32. Sipma WS, de Jong MFC, Meuleman Y, Hemmelder MH, Ahaus K. Facing the challenges of PROM implementation in Dutch dialysis care: patients' and professionals' perspectives. PLoS One. 2023;18(5):e0285822. [FREE Full text] [doi: 10.1371/journal.pone.0285822] [Medline: 37186606]

XSL•FO
RenderX

33. Saly E, Simon J, Fischer C. Healthcare professionals' perspectives on assessing selected patient-reported outcome measures in specialist palliative care institutions: a multi-country mixed-methods study. BMC Palliat Care. 2025;24(1):135. [FREE Full text] [doi: 10.1186/s12904-025-01775-6] [Medline: 40369516]
34. Amat-Fernandez C, Pardo Y, Ferrer M, Bosch G, Lizano-Barrantes C, Briseño-Diaz R, et al. Hospital de Mar Patient-Reported Measures Group. Evaluating the implementation of PROMs and PREMs in routine clinical care: co-design of tools from the perspective of patients and healthcare professionals. Health Qual Life Outcomes. 2025;23(1):15. [FREE Full text] [doi: 10.1186/s12955-025-02333-7] [Medline: 39962425]
35. Zhang YM, Wang YY, Liu Y, Zhong CC, Zhou M, Chen L, et al. Nurses' perspectives on the implementation of patient-reported outcomes: a descriptive qualitative study. BMC Nurs. 2025;24(1):1186. [FREE Full text] [doi: 10.1186/s12912-025-03847-7] [Medline: 41013401]

Abbreviations

CD: Crohn disease
COS: core outcome set
DM: diabetes mellitus
EHR: electronic health record
ePROM: electronic patient-reported outcome measure
FS: feasibility study
GRAMMS: Good Reporting of a Mixed Methods Study
H2O: Health Outcomes Observatory
HCP: health care professional
LC: lung cancer
MBC: metastatic breast cancer
PRO: patient-reported outcome
PROM: patient-reported outcome measure
PSQ-18: Patient Satisfaction Questionnaire Short Form
SUS: System Usability Scale
UC: ulcerative colitis
VBHC: value-based health care

Edited by A Mavragani; submitted 06.Mar.2026; peer-reviewed by M Bass, C Annema; comments to author 06.Aug.2026; revised version received 22.Sep.2026; accepted 24.Sep.2026; published 11.Nov.2026

Please cite as:
Long P, Galan G, Watson CE, Cossio-Gil Y, Andrews MR, Ritschl V, Schmolik V, Ferri M, Porth A-K, Otto LR, Rogge AA, Haberbosch L, Neureiter di Torrero A, Verbeij L, Serra-Ruiz X, Flores V, Far Serrano M, Lalji R, Primas C, Novacek G, Stamm T
Integrating Patient-Reported Outcome Measures Through the Health Outcomes Observatory (H2O) Digital Infrastructure: Multicenter Mixed Methods Feasibility Study
JMIR Form Res 2026;10:e94733
URL: https://formative.jmir.org/2026/1/e94733
doi: 10.2196/94733
PMID:

©Preston Long, Gemma Galan, Carolina E Watson, Yolima Cossio-Gil, Margaret Renn Andrews, Valentin Ritschl, Valentina Schmolik, Marta Ferri, Ann-Kristin Porth, Lisa Rebecca Otto, Alizé A Rogge, Linus Haberbosch, Anouk Neureiter di Torrero, Laura Verbeij, Xavier Serra-Ruiz, Vanesa Flores, Marta Far Serrano, Rahim Lalji, Christian Primas, Gottfried Novacek, Tanja Stamm. Originally published in JMIR Formative Research (https://formative.jmir.org), 11.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX