

Artificial Intelligence (AI)-powered Expertise: a real asset for identifying patient-reported measures (PROMs and PREMs) in clinical studies.

Jessica Soyer, Akram Hecini, Sylvain Juchet, Céline Desvignes-Gleizes, Maxime Thiebaut, Jean-Philippe Bertocchio

Submitted to: Journal of Medical Internet Research
on: September 21, 2025

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript.....	5
Supplementary Files.....	28
Multimedia Appendixes	29
Multimedia Appendix 1.....	29

Preprint
JMIR Publications

Artificial Intelligence (AI)-powered Expertise: a real asset for identifying patient-reported measures (PROMs and PREMs) in clinical studies.

Jessica Soyer^{1*} MSc; Akram Hecini^{1*} MSc; Sylvain Juchet¹ MSc; Céline Desvignes-Gleizes² PhD; Maxime Thiebaut¹ MSc; Jean-Philippe Bertocchio¹ MD, PhD

¹ SKEZI Annecy FR

² Mapi Research Trust Lyon FR

*these authors contributed equally

Corresponding Author:

Jean-Philippe Bertocchio MD, PhD

SKEZI

Les Papeteries - Image Factory, 1 Esplanade Augustin Aussédat

Annecy

FR

Abstract

Background: Traditional healthcare models have evolved to increasingly recognize patients perspectives as key to improving the quality of care, especially in oncology. In such, Patient-Centered Measures (PCMs), including Patient-Reported Outcome Measures (PROMs) and Patient-Reported Experience Measures (PREMs), can enhance Patient-Healthcare Provider (HCP) communication while facilitating individualized care. This tailored approach not only improves patient outcomes but also underscores the importance of research methodologies that actively account for variability in patient experiences, especially across different sociodemographic and clinical backgrounds. However, identifying PCMs in clinical research has traditionally been time-consuming.

Objective: We hypothesized that Artificial intelligence (AI) could automate and accelerate this identification. We conducted a study that aimed to estimate the proportion of clinical studies that included PROMs/PREMs, using either a traditional expert-based identification method or an AI-enriched approach.

Methods: In a retrospective cross-sectional study using the ClinicalTrial.gov database, we focused on oncology studies between 2012 and 2022. Two methods were assessed for identifying PROMs/PREMs: i) a traditional expert-based method, where an algorithm identified PROMs/PREMs from a list of 346 oncology-specific PROMs/PREMs (extracted from the PROQOLID™ database, Mapi Research Trust) and/or 11 PROMs/PREMs-specific terms; and ii) an AI-enriched method using a Bidirectional Encoder Representations from Transformers (BERT) model, trained on 2,399 outcomes labeled by experts. To evaluate the performance of the algorithms, we compared the results returned by each method to the gold standard (expert decision). Each study was classified as reporting using (or not) at least one PROM/PREM and was described accordingly (logistic regression). To better identify which PROMs/PREMs were the most frequently used in clinical research, a Named Entity Recognition (NER) model was then used.

Results: A total of 24,491 studies were included. According to the traditional expert-based algorithm, 7,549 studies (31%) used at least one PROM/PREM, as compared to 8,029 studies (33%) identified by the AI-enriched algorithm, increasing from 2012 to 2022 (Chi-squared test, $p < 0.001$). With 90% accuracy, the AI-enriched algorithm outperformed (Chi-squared test, $p < 0.001$) the traditional algorithm (83%) in identifying PROMs/PREMs. Breast and digestive cancers accounted for nearly 50% of all oncology studies using PROMs/PREMs, with the EORTC QLQ-C30 being the most frequently used. As expected, phases 2 to 4 trials more frequently included PROMs/PREMs than preclinical or early-phase 1 studies (OR[95% CI]=1.8[1.1-2.8] for phase 2; 3.6[2.3-5.8] for phase 3; 2.6,[1.6-4.4] for phase 4). In observational studies, cross-sectional and prospective studies incorporated PROMs/PREMs three times more frequently than retrospective studies (OR[95% CI]=4.6[3.3-6.4] and 3.2[2.5-4.1], respectively).

Conclusions: Both approaches relied on human expertise; however, the one that involved training AI has proven to be more effective than the traditional method in identifying PROMs/PREMs in clinical studies. Future studies should explore the

applicability of this approach to a broader field of pathologies and assess its applicability to other additional open-access databases. Clinical Trial: NA

(JMIR Preprints 21/09/2025:84533)

DOI: <https://doi.org/10.2196/preprints.84533>

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible to all users.

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <https://preprints.jmir.org/preprint/84533>

No. Please do not make my accepted manuscript PDF available to anyone. I understand that if I later pay to participate in <https://preprints.jmir.org/preprint/84533>

Original Manuscript

Title

Artificial Intelligence (AI)-powered Expertise: a real asset for identifying patient-reported measures (PROMs and PREMs) in clinical studies.

Authors

Jessica Soyer^{1*}, MSc, Akram Hecini^{1*}, MSc, Sylvain Juchet¹, MSc, Céline Desvignes-Gleizes², PhD, Maxime Thiebaut¹, PharmD, MSc, Jean-Philippe Bertocchio^{1,3,4,5}, MD, PhD

* contributed equally to this work

Affiliations

¹ SKEZI, Les Papèteries – Image Factory, 1 Esplanade Augustin Aussédat, 74960 Annecy, France

² Mapi Research Trust, 27 rue de la Villette, 69003 Lyon, France

³ Thyroïde – Tumeurs Endocrines, Pitié-Salpêtrière Hospital, Assistance Publique-Hôpitaux de Paris, Boulevard de l'hôpital, 75013 Paris, France

⁴ Reference center for rare diseases related to calcium-phosphate metabolism disorders, Pitié-Salpêtrière Hospital, Assistance Publique-Hôpitaux de Paris, Boulevard de l'hôpital, 75013 Paris, France

⁵ Université Paris Cité, INSERM U1333 Santé Orale, FHU-DDS-net, Dental School,

Montrouge, France

Correspondance

Jean-Philippe Bertocchio: jean-philippe.bertocchio@skezi.eu

Contributions

JS performed analysis, wrote first draft of the manuscript, revised the manuscript.

AH performed analysis, wrote first draft of the manuscript, revised the manuscript.

SJ designed the study, revised the manuscript.

CDG provided the list of PROMs/PREMs-related terms using the PROQOLID™ (Mapi Research Trust) database.

MT performed the analysis and the data layout, wrote further drafts of the manuscript, revised the manuscript, wrote the appendix.

JPB designed the study, supervised the study, wrote the first draft of the manuscript, revised the manuscript.

Data availability

Data will be made available upon reasonable request to the corresponding author.

Conflict of interest disclosure

SJ, HA, MT and JPB work at SKEZI, a digital platform that develops tools for collecting patient-reported outcomes. CDG works for a company that protects PROMs/PREMs.

Word counts 3,868

Abstract (457 words)

Background and objective. Traditional healthcare models have evolved to increasingly recognize patients perspectives as key to improving the quality of care, especially in oncology. In such, Patient-Centered Measures (PCMs), including Patient-Reported Outcome Measures (PROMs) and Patient-Reported Experience Measures (PREMs), can enhance Patient-Healthcare Provider (HCP) communication while facilitating individualized care. This tailored approach not only improves patient outcomes but also underscores the importance of research methodologies that actively account for variability in patient experiences, especially across different sociodemographic and clinical backgrounds. However, identifying PCMs in clinical research has traditionally been time-consuming. We hypothesized that Artificial intelligence (AI) could automate and accelerate this identification. We conducted a study that aimed to estimate the proportion of clinical studies that included PROMs/PREMs, using either a traditional expert-based identification method or an AI-enriched approach.

Material and Methods. In a retrospective cross-sectional study using the ClinicalTrials.gov database, we focused on oncology studies between 2012 and 2022. Two methods were assessed for identifying PROMs/PREMs: i) a traditional expert-based method, where an algorithm identified PROMs/PREMs from a list of 346 oncology-specific PROMs/PREMs (extracted from the PROQOLID™ database, Mapi Research Trust) and/or 11 PROMs/PREMs-specific terms; and ii) an AI-enriched method using a Bidirectional Encoder Representations from Transformers (BERT) model, trained on 2,399 outcomes labeled by experts. To evaluate the performance of the algorithms, we compared the results returned by each method to the gold standard (expert decision). Each study was classified as reporting using (or not) at least one PROM/PREM and was described accordingly (logistic regression). To better identify which PROMs/PREMs were the most frequently used in clinical research, a Named Entity Recognition (NER) model was then used.

Results. A total of 24,491 studies were included. According to the traditional expert-based algorithm, 7,549 studies (31%) used at least one PROM/PREM, as compared to 8,029 studies (33%) identified by the AI-enriched algorithm, increasing from 2012 to 2022 (Chi-squared test, $p < 0.001$). With 90% accuracy, the AI-enriched algorithm outperformed (Chi-squared test, $p < 0.001$) the traditional algorithm (83%) in identifying PROMs/PREMs. Breast and digestive cancers accounted for nearly 50% of all oncology studies using PROMs/PREMs, with the EORTC QLQ-C30 being the most frequently used. As expected, phases 2 to 4 trials more frequently included PROMs/PREMs than preclinical or early-phase 1 studies (OR[95% CI]=1.8[1.1-2.8] for phase 2; 3.6[2.3-5.8] for phase 3; 2.6,[1.6-4.4] for phase 4). In observational studies, cross-sectional and prospective studies incorporated PROMs/PREMs three times more frequently than retrospective studies (OR[95% CI]=4.6[3.3-6.4] and 3.2[2.5-4.1], respectively).

Conclusion. Both approaches relied on human expertise; however, the one that involved training AI has proven to be more effective than the traditional method in identifying PROMs/PREMs in clinical studies. Future studies should explore the applicability of this approach to a broader field of pathologies and assess its applicability to other additional open-access databases.

Trial Registration: NA

Keywords: Clinical trials; Patient-reported outcome measures; Patient-reported experience measures; Artificial intelligence-based research; Expert-based research; Oncology; Research on research.

Introduction

Traditional models in healthcare have evolved to increasingly recognize patients perspectives as key to improving their quality of services [1]. Data collected from Patient-Centered Measures (PCMs) made this shift possible [2]. PCMs refer to widely used and studied tools, such as Patients-Reported Outcome (PROMs) or experience (PREMs) measures [3], which constitute relevant information collected directly from patients to better understand and address what matters the most to them [2,4]. In their day-to-day practice, healthcare providers (HCP) may collect PCMs for various purposes such as screening, monitoring, evaluating care, or collecting side effects of treatments as well as to help in the decision-making process [5]. These measurements are added to biological measures and clinician-based outcomes, such as physical examination, to better plan patient care.

PROMs are standardized questionnaires that collect information about health outcomes directly from patients—such as symptoms, functional status, Health-Related Quality of Life (HRQoL), or mental well-being—without amendment or interpretation by a clinician or anyone else, and measure them in a quantitative manner at a specific timepoint. [6,7]. Originally developed in research studies, PROMs applications are now diversified and used in supporting clinical decision-making, assessing treatment benefit in drug trials, prioritizing patients for surgical procedures, comparing outcomes among HCPs, stimulating HRQoL improvement, and evaluating practices [8–10]. There are two main categories of PROMs: i) generic and ii) specific ones (including both condition-specific and concept-specific). Generic PROMs measure concepts relevant to a wide range of patients, enabling comparisons across different conditions. For example, the EQ-5D questionnaire [11,12] measures five dimensions of health: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Each dimension is assessed using a set of severity levels, initially comprising 3 levels (EQ-5D-3L) and expanded to 5 levels in the updated 2010 version (EQ-5D-5L), notably to improve sensitivity and reduce ceiling effect [12]. Specific PROMs, on the other hand, measure concepts relevant to a particular patient group or condition. For instance, the EORTC QLQ-C30 questionnaire measures HRQoL specifically in patients with cancer (condition-specific) [13], while the Fatigue Severity Scale (FSS) evaluates the concept of fatigue (concept-specific) [14]. Besides this core common questionnaire for all cancer patients, additional modules have been developed from it to measure symptoms and outcomes specific to different cancers. Examples include the EORTC Quality of Life Questionnaire - Lung Cancer Module (EORTC QLQ-LC13) or the Head and Neck Cancer Module (EORTC QLQ-H&N35) [15,16]. For the sole field of cancer, more than 300 PROMs have already been identified.

In contrast to PROMs, PREMs focus on patients experience, and emerge as a favored metric for assessing healthcare quality, such as the validated Kidney PREM questionnaire [17]. PREMs assess common information about patients satisfaction with service delivery in a clinical setting or describe patients experiences with a therapy or a plan of care [4]. Like the United-States Food and Drug Administration (FDA), other public health authorities in Europe, such as the Haute Autorité de Santé (HAS) in France, have recently added PROMs/PREMs to their list of care quality and safety indicators to encourage their use in clinical trials [18–20]. However, there is no established consensus regarding the way to name PROMs/PREMs. This makes their identification particularly challenging. New tools based on natural language processing (NLP) may be useful in this setting [21]. Overall, it appears essential to develop a solution for identifying and tracking the use of PROMs/PREMs in clinical research, both in terms of their frequency of inclusion in studies and the types of instruments selected. We hypothesized that an Artificial Intelligence-enriched algorithm would outperform an

expert-based algorithm in identifying PROMs/PREMs used in clinical research. Here, we report the results from a cross-sectional study that aimed to estimate the proportion of oncology clinical studies incorporating PROMs/PREMs and to examine the characteristics associated with these studies, using either a traditional expert-based method or an AI-enriched approach.



Material and Methods

Recruitment, data sources, and study design

To be included, studies had to be available in the open access ClinicalTrials.gov database, set in the oncology field, with a study start date between 01/01/2012 and 12/31/2022 and to contain at least one outcome described. Studies that contained no outcome description were therefore excluded.

Identifying PROMs/PREMs

To identify PROMs/PREMs, two different techniques were assessed: i) an expert-based algorithm and ii) an AI-enriched algorithm.

Expert-based algorithm

1- A list of 346 PROMs/PREMs used in oncology, including both their full names and acronyms, was extracted from the Patient-Reported Outcome and Quality of Life Instruments Database (PROQOLID™);

2- Moreover, two distinct groups of experts set up a list of terms related to PROMs/PREMs for the purpose of the present study. The experts then reached a consensus, resulting in a final list of PROMs/PREMs-related generic open terms. PROM-specific terms were: *quality of life*, *qol*, *patient-reported outcome*, *psychometric*, *patient perspective*, *questionnaire*, *health related quality of life*, and *hrql*. PREM-specific terms were: *satisfaction*, *experience*, and *quality of care*;

3- Using the ctrialsgov R package, studies were extracted from ClinicalTrials.gov.

4- Based on the expert-based algorithm, PROMs/PREMs and specific terms were specifically searched within the outcome descriptions available in the dataset.

5- To assess the effectiveness of this method, 1% of the identified studies were randomly selected. Pairs of experts independently reviewed each selected study to verify whether the PROMs/PREMs identified by this method were present and relevant.

AI-enriched algorithm

1- Data Collection: Using Bidirectional Encoder Representations from Transformers (BERT) [22], a text classification model was built. Texts from outcomes descriptions that were already labeled by pairs from our expert team as containing or not PROMs/PREMs were used. They were then segmented into three distinct subsets: a training, a validation, and a test subset, representing 60%, 13%, and 27% of the corpus, respectively; Finally, 2,399 outcomes were used for training, 479 for validation, and 1,065 for performance calculation;

2- Data encoding: using BERT Tokenizer [22], each word or textual element was converted into a token, which was then encoded as a fixed-length numerical vector of 512. This numerical vector captured the semantic and contextual information within the texts;

3- Training: Model training was conducted using the Trainer from the Transformers library developed by Hugging Face [23]. During training, multiple models were generated by adapting batch size and learning rate to get the best performance;

3- Testing and Evaluation: Every generated models were evaluated using the associated test subset. The performance of each model was assessed using the following metrics: accuracy, recall, and F1 score. These metrics gauged the model ability to correctly

classify texts into their respective categories. The best model was selected based on its accuracy score.

Evaluating the performances of algorithms

To evaluate their performances (accuracy, sensitivity, specificity, and precision), the results obtained by each algorithm were compared to a dataset of outcomes labeled manually by experts (reference set). Eight pairs of experts were randomly assigned a subset of outcomes to label. Each member of the pair did not know who their partner was and did not know what label was given by the other member of the pair. Outcomes were sampled to balance the number of positive and negative outcomes to be checked between the eight pairs: 2,570 positive outcomes (*i.e.* 15% of total positive outcomes) and 2,228 negative outcomes (*i.e.* 2% of total negative outcomes), as initially determined by the expert-based algorithm. Thus, a total of 4,798 outcomes through 892 studies were labeled by the 8 pairs of experts. Finally, the pairs met to reach a consensus in case of disagreement.

Describing the studies using PROMs/PREMs

Each study was classified as using at least one PROM/PREM or not. As observational and interventional studies carried out different characteristics, the analyses were stratified according to the type of study. We used multivariate logistic regression analysis to investigate potential associations between the use of PROMs/PREMs adjusted to other characteristics: cancer site, intervention type, clinical trial phase, randomization type, model type, and time perspective.

Recoding cancer sites

The ClinicalTrials.gov database returns a wide variety of namings for cancer sites, even for the same cancers, as this field is an open-label one. Cancer sites had to be recoded in order to enhance the homogeneity and accuracy of the data. First, a k-means clustering [24] was used to group the cancer sites based on the similarities between namings for each clinical study. The 10 most recurrent words in each cluster were identified for naming each cluster. The cluster names (cancer sites) were then used in a subsequent few-shot classification task, using the setfit model from Hugging Face [25]. For each cancer site, eight representative sentences were carefully selected from the dataset, which were used to create a final set of 116 texts. This dataset was used to train a text classification model, which was then applied to classify all the cancer sites in the dataset. The final model presented an accuracy of 96% for classifying cancer sites.

Identifying the most frequently used PROMs/PREMs

Although the main AI-enriched algorithm was designed to detect if a PROM/PREM was used or not, it was not designed to state which PROM/PREM it was referring to. To identify which very PROMs/PREMs were the most frequently used in clinical research, we used a Named Entity Recognition (NER) model called SKEP'IA®, that we had previously developed and patented. This model can recognize any PROM/PREM within a text, whether they are presented in full text or as an acronym and can suggest relevant ones based on textual descriptions. SKEP'IA® was developed by using labeled data collected from both ClinicalTrials.gov and PubMed. However, the absence of a standardized and official naming system for each PROM/PREM presented a challenge in determining the most commonly used ones, representing a limitation in the exclusive use of the SKEP'IA model. To address this issue, we implemented a text similarity

search approach using the PROQOLID™ database. Thus, the PROMs/PREMs identified by SKEP'IA® were compared to the list of PROMs/PREMs previously captured from the PROQOLID™ database by calculating text similarity between them. Finally, we combined both the NER model and the text similarity search in the PROQOLID™ database.

Statistical analysis

Statistical analyses were performed using R studio software, version 4.4.2. Descriptive statistics were used to illustrate the use of PROMs/PREMs in registered clinical studies. Characteristics of the studies were presented by their numbers and proportions. Proportions were compared bivariately using a Chi-square test. Year-over-year trends in the use of PROMs/PREMs were described. A multivariate logistic regression analysis was used to investigate possible associations between the use of PROMs/PREMs and study characteristics, by subgroups of interventional or observational studies. The variables included in the models were those obtained from the Clinicaltrial.gov database and considered relevant to the study. The Odds Ratios (OR) were presented using forest-plots, and statistical significance for each predictor was assessed using Wald tests. A p -value <0.05 was considered as reflecting statistical significance.

Results

From the selection of studies to PROMs/PREMs identification

Between 01/01/2012 and 12/31/2022, a total of 28,034 studies in oncology were found in the ClinicalTrials.gov database, including 24,491 with at least one outcome described (Figure 1A). According to the expert-based algorithm, 16,954 outcomes using PROMs/PREMs were found from 7,549 studies, meaning that out of the 24,491 eligible oncology studies, 31% used at least one PROM/PREM (Figure 1B). It exerted an 83% accuracy, an 86% sensitivity, and an 81% specificity (Table 1). The AI-enriched algorithm identified 8,029 studies that used at least one PROM/PREM, representing 33% (Figure 1C). It exerted a 90% accuracy, a 91% sensitivity, and an 89% specificity (Table 2). Therefore, the AI-enriched algorithm performed better (Chi-squared test, $p < 0.001$) than the expert-based algorithm in identifying PROMs/PREMs in oncology studies. The following results present the characteristic of studies identified based on the AI-enriched algorithm only. The results obtained using the expert-based algorithm are available in Supplementary Data.

Trends in PROMs/PREMs use in oncology studies overtime

The analysis of year-over-year trends revealed a steady increase in the proportion of clinical studies incorporating one or more PROMs/PREMs, from 27.6% in 2012 to 36.2% in 2021 (Figure 2).

Characteristics of studies

Table 3 summarizes the characteristics of the 24,491 studies included in the analysis, stratified on whether they used PROMs/PREMs ($n=8,029$) or not ($n=16,462$). Among all the included studies, 79% ($n=19,344$) were interventional trials. Focusing on interventional trials, Phase 2 (27.0%, $n=5,228$) was the most common study stage, followed by Phase 1 (13.1%, $n=2,536$), Phase 3 (8.5%, $n=1,650$), and Phase 1/2 (7.9%, $n=1,528$) to report using PROMs/PREMs. However, missing data regarding trial phases accounted for 37.7% ($n=7,285$) of studies. The most frequent allocation type was randomization, accounting for 44.4% ($n=8,586$) of interventional studies. The predominant intervention models were parallel assignment and single-group assignment, as they represented 48.2% ($n=9,330$) and 44.5% ($n=8,613$) of analyzed studies, respectively. The most common primary purpose of the included studies was treatment, accounting for 66.2% of cases ($n=12,815$), followed by supportive care with 10.6% ($n=2,058$). Among observational studies, the most prevalent observational model was cohorts, representing 64.0% ($n=3,295$) of the studies included in the analysis. In addition, 72.3% ($n=3,295$) of observational studies were prospective, making this the most frequently used time perspective.

Among all studies included in the analysis, 32.8% ($n=8,029$) incorporated at least one PROM/PREM. Out of those, 84.7% ($n=6,800$) were interventional, while 15.3% ($n=1,229$) were observational. Among interventional studies, Phase 2 and Phase 3 trials incorporated mostly PROMs/PREMs, representing 19.5% ($n=1,323$) and 12% ($n=814$) of them, respectively. Nevertheless, 55.4% ($n=3,767$) of interventional studies did not provide any exploitable data on their clinical phases. Interventional studies incorporating PROMs/PREMs were mainly randomized (64.7%, $n=4,401$). The main interventional models were mainly parallel assignment and single-group assignment, as they respectively represented 65% ($n=4,420$) and 29.3% ($n=1,994$) of studies with PROMs/PREMs. Treatment (55.7%, $n=3,789$) and supportive care (24.1%, $n=1,638$) were the most primary purposes of interventional studies using PROMs/PREMs. Among observational studies using PROMs/PREMs ($n=1,229$), 67.9% ($n=835$) were cohort

studies, while 12.7% (n=156) were case-only studies. Considering time perspective, prospective studies were the most frequent to incorporate PROMs/PREMs, representing 79.8% (n=981) of them.

PROMs/PREMs use by sites of cancer

Proportions of organ-specific oncology studies using at least one PROM/PREM are shown in Figure 3. Breast and digestive cancers accounted for nearly 50% of all studies using PROMs/PREMs in oncology. Urogenital, lung, and prostate cancers followed, representing 8%, 10%, and 11%, respectively, of all studies using PROMs/PREMs in oncology.

PROMs and PREMs used in oncology studies

Among the 8,029 studies that included at least one PROM/PREM, 4,823 (61.1%) used one or more of the nine most commonly employed PROMs/PREMs. Among these, the EORTC QLQ-C30 was the most frequently used PROM/PREM in clinical studies in oncology. It was utilized in approximately 33.2% (n=1,602) studies included in our analysis. The FACT-G and the EQ-5D-5L questionnaires were the next most frequently used PROMs/PREMs, appearing in 15.2% (n=734) and 11.5% (n=552) of the studies analyzed, respectively (Table 4). Table 5 presents the categories of health dimensions assessed by the instruments identified across the 4,823 studies that used one or more of the nine most commonly employed PROMs/PREMs. In 40.7% (n=1,969) of the identified cases, the PROMs/PREMs focused on care-related patient satisfaction, followed by organ-specific quality of life (n=783; 16.2%), as well as depression and anxiety scales (n=441; 9.1%). This distribution results from the zero-shot, multi-label classification model, which often assigned satisfaction in overlap with quality of life. The latter was further divided into general and cancer-specific subcategories. No additional filters based on study type or trial phase were applied to obtain these proportions.

Figure 4 provides the distribution of the various scales commonly used in identified clinical studies, categorized by cancer site. Breast and digestive cancers were the most represented in the use of PROMs/PREMs in oncology studies, accounting for 22% and 20% of cases, respectively. These were followed by prostate (11%), lung (10%), urogenital (8%), and head and neck cancers (8%). In contrast, hematopoietic (3%), brain (3%), thyroid (1%), skin (1%), bone (1%), and neuroendocrine cancers (0.4%) were less frequently studied using PROMs/PREMs, each representing fewer than 5% of cases.

The use of PROMs/PREMs varied according to several characteristics of studies (Figure 5A). Out of interventional studies, phases 2 to 4 studies were incorporating more often PROMs/PREMs than pre-clinic or early phase 1 ones (OR[95%CI]=1.8[1.1-2.8] for phase 2; OR[95%CI]=3.6[2.3-5.8] for phase 3; OR[95%CI]=2.6[1.6-4.4] for phase 4). Randomized Clinical Trials (RCTs) used twice as many PROMs/PREMs than non-randomized trials (OR[95%CI]=2.0[1.7-2.4]). Clinical trials focusing on supportive care demonstrated significantly higher odds for using PROMs/PREMs than trials with other objectives (OR[95%CI]=4.1[2.7-6.2]). Studies concerning some organ-specific trials were less likely to use PROMs/PREMs than non-organ-specific ones – such as breast (OR[95% CI]=0.7[0.5-0.9]), digestive (OR[95% CI]=0.6[0.5-0.8]), lung (OR[95% CI]=0.7[0.4-0.7]), haemopoietic (OR[95% CI]=0.5[0.3-0.8]), and thyroid (OR[95% CI]=0.4[0.2-0.9]) cancers –. Similarly, the use of PROMs/PREMs instruments in observational studies differed according to characteristics of studies (Figure 5B). In contrast to retrospective studies, cross-sectional and prospective studies used more often (OR[95%CI]=4.6[3.3-6.4] and 3.2[2.5-4.1], respectively) PROMs/PREMs. After adjusting for confounding factors, cohort studies had significantly higher odds for using

PROMs/PREMs than case-control studies (OR[95%CI]=1.6[1.3-2.1]). Additionally, studies focusing on organ-specific trials — such as breast (-30%, OR[95% CI]=0.7[0.5-0.8]), digestive (OR[95% CI]=0.4[0.3-0.5]), head and neck (OR[95% CI]=0.6[0.4-0.8]), hematopoietic (OR[95% CI]=0.6[0.4-0.8]), lung (OR[95% CI]=0.4[0.3-0.5]), thyroid (OR[95% CI]=0.5[0.3-0.9]), and urogenital (OR[95% CI]=0.4[0.3-0.6]) cancers — had lower odds for using PROMs/PREMs than non-organ-specific studies.

Discussion

Taken together, our results clearly show that i) an AI-enriched algorithm outperformed an expert-based one in identifying PROMs/PREMs in clinical studies, ii) the use of PROMs/PREMs has increased over time, iii) studies in breast cancer and digestive cancer accounted for nearly half of the studies incorporating PROMs/PREMs, and iv) interventional prospective studies were more likely to use PROMs/PREMs than any other type of studies.

The AI-enriched algorithm that we developed for this specific purpose appears particularly promising for researchers seeking to identify both the presence of PROMs/PREMs and which PROM/PREM was used in oncology studies. With a 90% accuracy, we were able to better identify studies than experts would do manually and in a more efficient way: first, the manual checking of 4,798 outcomes (accounting for 892 studies) took several weeks using a work force of 16 high-skilled people while the algorithm was able to classify the whole 24,491 studies within a few minutes (from 10 minutes to 3 hours, depending on the GPU workforce). Although the development of the AI-enriched algorithm required a considerable initial investment in time and expertise, this effort represents a long-term return on investment, as the tool can subsequently process large datasets much faster than expert-based approaches. A significant increase in the use of PROMs/PREMs in phase 2 to 4 studies could be forecasted in the coming years, driven by the growing interest from key stakeholders, including industry, regulatory bodies, patients, and patient associations, as well as their commitment to promoting value-based healthcare (VBHC) in clinical practice. Several regulatory bodies, such as the FDA and HAS, have recently incorporated PROMs/PREMs into their care quality and safety indicators to encourage their use in clinical trials [18-20]. Additionally, the increasing incorporation of PROMs/PREMs in clinical studies that we report supports our optimism. The AI-enriched algorithm tool could be extended to other medical indications, fostering widespread research on PCM. On a larger scale, AI could become an invaluable tool for medical stakeholders to manage, conduct, and monitor clinical studies [26]. Nevertheless, the AI-driven method also carries the risk of misclassifying conceptually close types of clinical outcome assessments (e.g., observer-reported outcomes or clinician-reported outcomes) as PROMs/PREMs, potentially leading to an overestimation of their prevalence compared with the expert-based review. In contrast, the expert-based review offers the advantage of manual verification by trained reviewers, which helps ensure accuracy and contextual understanding, even if it is more resource-consuming.

Among all organ-specific studies, PROMs/PREMs were more frequently used in studies focusing on breast and digestive cancers. This trend may be attributed to the extensive research on these cancers, driven by their high prevalence, significant impact on patients quality of life, and the growing need for patient feedback as treatments evolve [27,28]. Furthermore, the diverse patient populations affected by breast and digestive cancers—spanning various age groups, genders, and ethnic backgrounds—underscore the importance of considering patients' perceived experiences and outcomes. While this is not unique to these cancer sites, breast and digestive cancers are among the most extensively studied in terms of patient-reported outcomes [29,30], likely due to their high prevalence and the strong presence of advocacy groups and research funding. This visibility has facilitated the integration of VBHC principles, where patient-centered care remains a central theme in medical research [31]. Compared to other cancer sites, breast and digestive cancers feature more prominently in VBHC-related literature, highlighting a gap that should be addressed for less-studied malignancies. It is also possible that in therapeutic areas where PROMs/PREMs are less common or less standardized, the relative performance gap between AI- and expert-based methods could be smaller, underscoring the importance of context when interpreting these

findings. In our dataset, where PROMs/PREMs were present in 33% of oncology studies, the algorithm achieved strong sensitivity and specificity; however, because sensitivity and specificity depend on the prevalence of the measured outcome, its performance may be reduced in contexts where PROMs/PREMs are less frequently or less consistently reported. Further studies are therefore needed to assess the robustness and generalizability of AI-driven approaches across diverse conditions.

We also report that interventional/prospective studies incorporated more PROMs/PREMs than observational/retrospective studies. This trend could be attributed to the growing interest of regulatory bodies and sponsors (e.g., industry, investigators, promoters) in incorporating these measures into clinical trials to support, particularly, regulatory, including regulators, like FDA and European Medicines Agency, and Health Technology Assessment (HTA) approvals. In such, some governments have begun recommending the use of PROMs/PREMs in interventional studies through official communications and guidelines [19,32]. Interventional studies are, by definition, designed with a structured protocol that allows for the systematic collection of specific data, making them particularly suited for incorporating PROMs/PREMs. Furthermore, since PROMs/PREMs provide direct patient insights, their inclusion in interventional studies is especially relevant, as these studies aim to evaluate the effects of specific treatments or interventions. Incorporating PROMs/PREMs into the design of observational studies requires their integration into routine care. Another approach would involve leveraging PCMs in registries or other real-world data sources and incentivizing the use of PROMs/PREMs in funding and in publishing by updating publication guidelines and funding requirements. Collaborative approaches, such as engaging patient advocacy groups and fostering multidisciplinary teams, may further enhance their adoption, as these groups often emphasize holistic, patient-centered care, and bring together diverse clinical perspectives that recognize the value of patient-reported measures.

The increase use of PROMs/PREMs in oncology was also earlier reported by Vodicka *et al.* (2015), which aimed to estimate the proportion of clinical trials using at least one PROM/PREM from November 2007 to December 2013 [33]. They identified 29% of oncology trials using PROMs/PREMs where we identified 33%. This could either be due to an improvement in the method for identifying studies or to an increase in the use of PROMs/PREMs over time. We used the same ClinicalTrials.gov database with a larger number of studies. The main difficulty in identifying PROMs/PREMs relied in the classification used in the ClinicalTrials.gov database. No standardized naming for PROMs/PREMs was currently available, leading researchers and healthcare providers to use the same PROMs/PREMs but with different namings. This observation may explain the poor quality of data retrieved from studies used by our AI-enriched algorithm. Both algorithms (AI-enriched and expert-based ones) exhibited the same issue, but our AI-enriched algorithm remained more efficient than the expert-based one. Therefore, classification appeared to be the major risk of bias in our study. Both the trends in the use of PROMs/PREMs in oncology trials over time and their distribution exhibited performance rates below 100%, due to misclassification of cancers. One potential way to reduce this bias would be the implementation of an AI-based tool to support researchers in identifying, labeling and naming PROs, thereby enhancing consistency in classification, terminology (names), and acronyms.

Overall, within the oncology field, evidence suggests that routine collection of PROMs/PREMs enhances patient-provider communication, helps identify symptoms and/or concerns, and improves patient satisfaction [34]. However, there is currently limited evidence demonstrating their impact in oncological settings, particularly regarding an effective improvement of quality or a better healthcare system performance [35,36]. Therefore, further research is essential to address these gaps, assess the applicability of this approach across a broader range of pathologies, and evaluate its adaptability to additional open-access databases. Such investigations would help determine the

generalizability and scalability of the methodology, facilitating its integration into diverse healthcare data systems and supporting its potential for a widespread application in health data analysis and patient-centered research.

Conclusion

Altogether, our data clearly demonstrate that an AI-enriched algorithm was more efficient (both more sensitive and more specific) as well as more efficient (time saving) than an expert-based algorithm in identifying PROMs/PREMs in oncology clinical studies. PROMs/PREMs are being used more and more over time, especially in interventional/prospective studies. We suggest that a better naming/classification of PROMs/PREMs could facilitate their identification and use, for example through the development of an AI-based tool to assist researchers in identifying, labeling and naming PROs. Further studies should extend this identification to other types of pathologies available in the ClinicalTrials.gov database and/or to other databases.

Acknowledgements

The authors thank all the physicians, researchers, and patients using PROMs/PREMs in clinical research, as well as all those who develop and validate patient-reported measures.

Conflicts of Interest

JS, AH, SJ, MT and JPB work at SKEZI, a company that develops tools for improving the use of patient-reported measures. CDG works at Mapi Research Trust, a company that licenses PROMs/PREMs.

Abbreviations

BERT: Bidirectional encoder representations from transformers

CI: Confidence Interval

EORTC: European Organization for Research and Treatment of Cancer

EQ-5D-3L: EuroQoL 5-Dimension 3-Level

EQ-5D-5L: EuroQoL 5-Dimension 5-Level

FDA: Food and Drug Administration

HAS: *Haute Autorité de Santé* - French Authority of Health

NER: Named entity recognition

NIH: National Institutes of Health

OR: Odds Ratio

PCM: Patient-centered measures

PREs: Patient-reported experiences

PREMs: Patient-reported experience measures

PROMs: Patient-reported outcomes measures

PROQOLID™: Patient-Reported Outcome and Quality of Life Instruments Database

QLQ: Quality of Life Questionnaire

QLQ-C30: Quality of Life Questionnaire-Core 30

References

1. Marshall S, Haywood K, Fitzpatrick R. Impact of patient-reported outcome measures on routine practice: a structured review. *J Eval Clin Pract* 2006 Oct;12(5):559–568. PMID:16987118
2. Howard AF, Warner L, Cuthbertson L, Sawatzky R. Patient-driven research priorities for patient-centered measurement. *BMC Health Serv Res* 2024 Jun 14;24(1):735. PMID:38877556
3. Churrua K, Pomare C, Ellis LA, Long JC, Henderson SB, Murphy LED, Leahy CJ, Braithwaite J. Patient-reported outcome measures (PROMs): A review of generic and condition-specific measures and a discussion of trends and issues. *Health Expect* 2021 Aug;24(4):1015–1024. PMID:33949755
4. Wolff AC, Dresselhuis A, Hejazi S, Dixon D, Gibson D, Howard AF, Liva S, Astle B, Reimer-Kirkham S, Noonan VK, Edwards L. Healthcare provider characteristics that influence the implementation of individual-level patient-centered outcome measure (PROM) and patient-reported experience measure (PREM) data across practice settings: a protocol for a mixed methods systematic review with a narrative synthesis. *Syst Rev* 2021 Jun 9;10(1):169. PMID:34108024
5. Sawatzky R, Laforest E, Schick-Makaroff K, Stajduhar K, Reimer-Kirkham S, Krawczyk M, Öhlén J, McLeod B, Hilliard N, Tayler C, Robin Cohen S. Design and introduction of a quality of life assessment and practice support system: perspectives from palliative care settings. *J Patient Rep Outcomes* 2017;2:36. PMID:30175318
6. Lavalley DC, Chenok KE, Love RM, Petersen C, Holve E, Segal CD, Franklin PD. Incorporating Patient-Reported Outcomes Into Health Care To Engage Patients And Enhance Care. *Health Aff (Millwood)* 2016 Apr;35(4):575–582. PMID:27044954
7. FDA-NIH Biomarker Working Group. BEST (Biomarkers, EndpointS, and other Tools) Resource. Silver Spring (MD): Food and Drug Administration (US); 2016. PMID:27010052
8. Dawson J, Doll H, Fitzpatrick R, Jenkinson C, Carr AJ. The routine use of patient reported outcome measures in healthcare settings. *BMJ* 2010 Jan 18;340:c186. PMID:20083546
9. Chen J, Ou L, Hollis SJ. A systematic review of the impact of routine collection of patient reported outcome measures on patients, providers and health organisations in an oncologic setting. *BMC Health Serv Res* 2013 Jun 11;13:211. PMID:23758898
10. Price A, Smith J, Dakin H, Kang S, Eibich P, Cook J, Gray A, Harris K, Middleton R, Gibbons E, Benedetto E, Smith S, Dawson J, Fitzpatrick R, Sayers A, Miller L, Marques E, Gooberman-Hill R, Blom A, Judge A, Arden N, Murray D, Glyn-Jones S, Barker K, Carr A, Beard D. The Arthroplasty Candidacy Help Engine tool to select candidates for hip and knee replacement surgery: development and economic modelling. *Health Technol Assess* 2019 Jun;23(32):1–216. PMID:31287051
11. van Hout B, Janssen MF, Feng Y-S, Kohlmann T, Busschbach J, Golicki D, Lloyd A, Scalone L, Kind P, Pickard AS. Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. *Value Health* 2012;15(5):708–715. PMID:22867780
12. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, Bonsel G, Badia X. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res* 2011 Dec;20(10):1727–1736. PMID:21479777
13. Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, Filiberti A, Flechtner H, Fleishman SB, de Haes JC. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *J Natl Cancer Inst* 1993 Mar 3;85(5):365–376. PMID:8433390
14. Ruotolo I, Carenza A, Sellitto G, Simeon R, Panuccio F, Marini Padovani A, Amadio E, Ugolini A, Berardi A, Galeoto G. Exploring the psychometric properties of the fatigue severity scale: results from a systematic review and reliability meta-analysis. *Expert Rev Pharmacoecon Outcomes Res* 2025 Sep;25(7):1021–1036. PMID:40408308

15. Bergman B, Aaronson NK, Ahmedzai S, Kaasa S, Sullivan M. The EORTC QLQ-LC13: a modular supplement to the EORTC Core Quality of Life Questionnaire (QLQ-C30) for use in lung cancer clinical trials. EORTC Study Group on Quality of Life. *Eur J Cancer* 1994;30A(5):635–642. PMID:8080679
16. Noel CW, Stephens RF, Su JS, Xu W, Krahn M, Monteiro E, Goldstein DP, Giuliani M, Hansen AR, de Almeida JR. Mapping the EORTC QLQ-C30 and QLQ-H&N35, onto EQ-5D-5L and HUI-3 indices in patients with head and neck cancer. *Head Neck* 2020 Sep;42(9):2277–2286. PMID:32333829
17. Hawkins J, Smeeton N, Busby A, Wellsted D, Rider B, Jones J, Steenkamp R, Stannard C, Gair R, van der Veer SN, Corps C, Farrington K. Contributions of treatment centre and patient characteristics to patient-reported experience of haemodialysis: a national cross-sectional study. *BMJ Open* 2021 Apr 14;11(4):e044984. PMID:33853800
18. Leandre. Aide à l'utilisation de questionnaires patients de mesure des résultats de soins dans le cadre de l'expérimentation «Épisode de soins». Saint-Denis-La-Plaine: Haute Autorité de santé (HAS) 2019;
19. Haute Autorité de Santé. Quality of care as perceived by patients – PROMs and PREMs indicators Panorama of experiences in other countries and main lessons learned. 2021. Available from: https://www.has-sante.fr/upload/docs/application/pdf/2022-03/proms_prem_panorama_of_experiences_in_other_countries_vf.pdf
20. U.S. Department of Health and Human Services FDA Center for Drug Evaluation and Research, U.S. Department of Health and Human Services FDA Center for Biologics Evaluation and Research, U.S. Department of Health and Human Services FDA Center for Devices and Radiological Health. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. *Health Qual Life Outcomes* 2006 Oct 11;4:79. PMID:17034633
21. Lu Z, Sim J-A, Wang JX, Forrest CB, Krull KR, Srivastava D, Hudson MM, Robison LL, Baker JN, Huang I-C. Natural Language Processing and Machine Learning Methods to Characterize Unstructured Patient-Reported Outcomes: Validation Study. *J Med Internet Res* 2021 Nov 3;23(11):e26777. PMID:34730546
22. Devlin J, Chang M-W, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv.org*. 2018. Available from: <https://arxiv.org/abs/1810.04805v2> [accessed Apr 2, 2025]
23. Wolf T, Debut L, Sanh V, Chaumond J, Delangue C, Moi A, Cistac P, Rault T, Louf R, Funtowicz M, Davison J, Shleifer S, Platen P von, Ma C, Jernite Y, Plu J, Xu C, Scao TL, Gugger S, Drame M, Lhoest Q, Rush AM. HuggingFace's Transformers: State-of-the-art Natural Language Processing. *arXiv*; 2020. doi: 10.48550/arXiv.1910.03771
24. Hartigan JA, Wong MA. Algorithm AS 136: A K-Means Clustering Algorithm. *Journal of the Royal Statistical Society Series C (Applied Statistics)* [Royal Statistical Society, Oxford University Press]; 1979;28(1):100–108. doi: 10.2307/2346830
25. Tunstall L, Reimers N, Jo UES, Bates L, Korat D, Wasserblat M, Pereg O. Efficient Few-Shot Learning Without Prompts. *arXiv*; 2022. doi: 10.48550/arXiv.2209.11055
26. Salvagno M, Cassai AD, Zorzi S, Zaccarelli M, Pasetto M, Sterchele ED, Chumachenko D, Gerli AG, Azamfirei R, Taccone FS. The state of artificial intelligence in medical research: A survey of corresponding authors from top medical journals. *PLoS One* 2024;19(8):e0309208. PMID:39178224
27. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, Jemal A. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2024;74(3):229–263. PMID:38572751
28. Arnold M, Abnet CC, Neale RE, Vignat J, Giovannucci EL, McGlynn KA, Bray F. Global

Burden of 5 Major Types Of Gastrointestinal Cancer. *Gastroenterology* 2020 Jul;159(1):335-349.e15. PMID:32247694

29. Rutherford C, Campbell R, White K, King M. Patient-reported outcomes as predictors of survival in patients with bowel cancer: a systematic review. *Qual Life Res* 2019 Nov 1;28(11):2871–2887. doi: 10.1007/s11136-019-02255-0

30. Salmani H, Nasiri S, Alemrajabi M, Ahmadi M. Advancing patient-centered cancer care: a systematic review of electronic patient-reported outcome measures. *Front Rehabil Sci Frontiers*; 2024 Sep 25;5. doi: 10.3389/fresc.2024.1427712

31. Demedts I, Himpe U, Bossuyt J, Anthoons G, Bode H, Bouckaert B, Carron K, Dobbelaere S, Mariën H, Van Haecke P, Verbeke W. Clinical implementation of value based healthcare: Impact on outcomes for lung cancer patients. *Lung Cancer* 2021 Dec 1;162:90–95. doi: 10.1016/j.lungcan.2021.10.010

32. U.S. Department of Health and Human Services FDA Center for Drug Evaluation and Research, U.S. Department of Health and Human Services FDA Center for Biologics Evaluation and Research, U.S. Department of Health and Human Services FDA Center for Devices and Radiological Health. Guidance for Industry: Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. 2009; Available from: <https://www.fda.gov/media/77832/download>

33. Vodicka E, Kim K, Devine EB, Gnanasakthy A, Scoggins JF, Patrick DL. Inclusion of patient-reported outcome measures in registered clinical trials: Evidence from ClinicalTrials.gov (2007-2013). *Contemp Clin Trials* 2015 Jul;43:1–9. PMID:25896116

34. Graupner C, Kimman ML, Mul S, Slok AHM, Claessens D, Kleijnen J, Dirksen CD, Breukink SO. Patient outcomes, patient experiences and process indicators associated with the routine use of patient-reported outcome measures (PROMs) in cancer care: a systematic review. *Support Care Cancer* 2021;29(2):573–593. PMID:32875373

35. Chen J, Ou L, Hollis SJ. A systematic review of the impact of routine collection of patient reported outcome measures on patients, providers and health organisations in an oncologic setting. *BMC Health Serv Res* 2013 Jun 11;13:211. PMID:23758898

36. Howell D, Molloy S, Wilkinson K, Green E, Orchard K, Wang K, Liberty J. Patient-reported outcomes in routine cancer clinical practice: a scoping review of use, impact on health outcomes, and implementation factors. *Ann Oncol* 2015 Sep;26(9):1846–1858. PMID:25888610

Tables

Table 1. Confusion matrix based on outcomes determined by experts-based method

		<i>Observed values</i>	
		Yes	No
<i>Predicted values</i>	Use of PROMs/PREMs		
	Yes	1 684	539
	No	264	2 311

Legend: The table shows the classification performance of the expert-based model predicting the use of Patient-Reported Outcome Measures (PROMs) and Patient-Reported Experience Measures (PREMs). "Yes" and "No" refer to whether PROMs/PREMs were used, as determined by experts. The model correctly predicted use in 1,684 cases and non-use in 2,311 cases, with 539 false positives and 264 false negatives.

Table 2. Confusion matrix based on outcomes determined by the AI-enriched algorithm

		<i>Observed values</i>	
		Yes	No
<i>Predicted values</i>	Use of PROMs/PREMs Yes	460	60
	No	46	499

Legend: This table displays the performance of the AI-enriched model for identifying the use of Patient-Reported Outcome Measures (PROMs) and Patient-Reported Experience Measures (PREMs). The model correctly predicted usage in 460 cases and non-usage in 499 cases, with 60 false positives and 46 false negatives.

Table 3. Description of the characteristics of studies according to their use of PROMs/PREMs (AI-enriched algorithm).

CHARACTERISTICS	All N= 24,491	Studies not using a PROM/PREM N= 16,462	Studies that used PROM/ PREM N= 8,029	p- value
Study type				<
Interventional	19,344 (79.0%)	12,544 (76.2%)	6,800 (84.7%)	0.001
Observational	5,147 (21.0%)	3,918 (23.8%)	1,229 (15.3%)	
INTERVENTIONAL STUDIES	19,344	12,544	6,800	
Clinical Phase				<
Early Phase 1	376 (1.9%)			0.001 ^a
Phase 1	2,536 (13.1%)			
Phase 1/Phase 2		314 (2.5%)	62 (0.9%)	
Phase 2	1,528 (7.9%)	2,253 (18.0%)	283 (4.2%)	
Phase 2/Phase 3		1,279 (10.2%)	249 (3.7%)	
Phase 3	5,228 (27.0%)	3,905 (31.1%)	1,323 (19.5%)	
Phase 4		165 (1.3%)	124 (1.8%)	
Missing ¹	289 (1.5%)	836 (6.7%)	814 (12.0%)	
	1,650 (8.5%)	274 (2.2%)	178 (2.6%)	
	452 (2.3%)	3,518 (28.0%)	3,767 (55.4%)	
	7,285 (37.7%)			
Allocation type				<
Non-Randomized	2,543 (13.1%)			0.001 ^a
Randomized		2,025 (16.1%)	518 (7.6%)	
NA ²	8,586 (44.4%)	4,185 (33.4%)	4,401 (64.7%)	
	8,215 (42.5%)	6,334 (50.5%)	1,881 (27.7%)	
Interventional model				<
Crossover Assignment	374 (1.9%)			0.001 ^a
Factorial Assignment	120 (0.6%)	206 (1.6%)	168 (2.5%)	
Parallel Assignment	9,330 (48.2%)	58 (0.5%)	62 (0.9%)	
Sequential Assignment		4,910 (39.1%)	4,420 (65.0%)	
Single Group Assignment	873 (4.5%)	725 (5.8%)	148 (2.2%)	
Missing	8,613 (44.5%)	6,619 (52.8%)	1,994 (29.3%)	
	34 (0.2%)	26 (0.2%)	8 (0.1%)	
Primary purpose				<
Diagnostic	1,573 (8.1%)			0.001 ^a
Prevention				
Supportive Care	1,031 (5.3%)	1,401 (11.1%)	172 (2.5%)	
Treatment		561 (4.5%)	470 (6.9%)	
Other	2,058 (10.6%)	420 (3.3%)	1,638 (24.1%)	
Missing		9,026 (72.0%)	3,789 (55.7%)	
	12,815 (66.2%)	1,082 (8.6%)	705 (10.4%)	
	1,787 (9.2%)	54 (0.4%)	26 (0.4%)	
	80 (0.4%)			
OBSERVATIONAL STUDIES	5,147	3,918	1,229	
Observational model				<
Case-Control	535 (10.4%)			0.001 ^a
Case-Only				
Cohort	736 (14.3%)	441 (11.3%)	94 (7.6%)	
Other		580 (14.8%)	156 (12.7%)	
Missing	3,295 (64.0%)	2,460 (62.8%)	835 (67.9%)	
		396 (10.1%)	125 (10.2%)	
	521 (10.1%)	41 (1.0%)	19 (1.5%)	
	60 (1.2%)			

Time Perspective

Cross-Sectional	359 (7.0%)			< 0.001 ^a
Prospective	3,723	243 (6.2%)	116 (9.4%)	
Retrospective	(72.3%)	2,742 (70.0%)	981 (79.8%)	
Other	836	753 (19.2%)	83 (6.8%)	
Missing	(16.2%)	161 (4.1%)	43 (3.5%)	
	204 (4.0%)	19 (0.5%)	6 (0.5%)	
	25 (0.5%)			

N (%)

Abbreviations: AI: Artificial intelligence; NIH: National Institute of Health; PREMs: Patient-Reported Experience Measures; PROMs: Patient-Reported Outcomes Measures

Legend: This table presents the distribution of key study characteristics for all included studies (N = 24,491), stratified by whether they used Patient-Reported Outcome Measures (PROMs) or Patient-Reported Experience Measures (PREMs). Characteristics are shown separately for interventional and observational studies, with proportions compared between studies using and not using PROMs/PREMs. Significant differences were found across study type, clinical phase, allocation method, interventional model, primary purpose, observational model, and time perspective (all $p < 0.001$).

Details:

^a : p-values do not account for missing data, as the data are censored.

¹ : Missing data does not correspond to any clinical trial phase as per its definition (e.g., medical devices).

² : Trials classified as Single Group Assignment and Sequential Assignment are grouped under "NA" for allocation type.

Table 4. Instruments repartition among studies that used one or more of the nine most commonly employed PROMs/PREMs (n=4,823)

Instrument	Value	Proportion (%)
EORTC QLQ-30	1,602	33.22
FACT-G	734	15.22
Cancer-Related QoL	416	8.63
HADS	441	9.14
C-VAS	263	5.45
NCI	411	8.52
EQ-5D-5L	552	11.45
PRO-CTCAE	292	6.05
IPSS	112	2.32
Total	4,823	100.00

Abbreviations: Cancer Related QoL: Cancer-Related Quality of Life ; C-VAS: Cancer Visual Analogue Scale ; EORTC QLQ-C30: European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire – Core 30 ; EQ-5D-5L: EuroQol 5 Dimensions – 5 Levels ; FACT-G: Functional Assessment of Cancer Therapy – General ; HADS: Hospital Anxiety and Depression Scale ; IPSS: International Prostate Symptom Score ; NCI: National Cancer Institute ; PREMs: Patient-Reported Experience Measures ; PRO-CTCAE: Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events ; PROMs: Patient-reported Outcomes Measures

Legend: This table summarizes the frequency and proportion of studies that employed at least one of the nine most frequently used Patient-Reported Outcome Measures (PROMs) or Patient-Reported Experience Measures (PREMs). The most commonly used instrument was the EORTC QLQ-C30 (33.22%), followed by FACT-G (15.22%) and EQ-5D-5L (11.45%). Less frequently used instruments included IPSS and C-VAS.

Table 5. Categories of Health Dimensions Assessed by Instruments Identified among studies that used one or more of the nine most commonly employed PROMs/PREMs (n=4,823)

Category	Value	Proportion (%)
Patient satisfaction with care	1,969	40.74
Organ-specific quality of life	783	16.20
Depression and anxiety scale	441	9.12
NA	1,640	33.93
Total	4,823	100.00

Legend: This table shows the distribution of health dimensions evaluated by instruments used in studies employing one or more of the nine most frequently used Patient-Reported Outcome Measures (PROMs) and Patient-Reported Experience Measures (PREMs). The majority of instruments assessed patient satisfaction with care (40.74%), followed by organ-specific quality of life (16.20%) and depression and anxiety scales (9.12%). A notable proportion (33.93%) of assessments were classified as not applicable (NA).

Supplementary Files

Multimedia Appendixes

Untitled.

URL: <http://asset.jmir.pub/assets/69a687ce4e7d1cfac9f0505e0b062098.docx>