Standardising the assessment of patient-reported outcome measures in localised prostate cancer. A systematic review

PIONEER Consortium; Ratti, Maria; Gandaglia, Giorgio; Alleva, Eugenia; Leardini, Luca; Sisca, Elena; Derevianko, Alexandra; Furnari, Federica; Ferracini, Serena; Beyer, Katharina; Moss, Charlotte; Pellegrino, Francesco; Sorce, Gabriele; Barletta, Francesco; Scuderic, Simone; Omar, Muhammad Imran; MacLennan, Steven; Williamson, Paula R.; Zong, Jihong; MacLennan, Sara J.

Standardizing the Assessment of Patient-Reported Outcome Measures in Localised Prostate Cancer. A Systematic Review

Maria Monica Ratti[a,b*], Giorgio Gandaglia[c*], Eugenia Alleva[a], Luca Leardini[b], Elena Silvia Sisca[b], Alexandra Derevianko[b], Federica Furnari[b], Serena Mazzoleni Ferracini[b], Katharina Beyer[d], Charlotte Moss[d], Francesco Pellegrino[c], Gabriele Sorce[c], Francesco Barletta[c], Simone Scuderi[c], Muhammad Imran Omar[e], Steven MacLennan[e], Paula R. Williamson[f], Jihong Zong[g]; Sara J MacLennan[e], Nicolas Motteth[h], Philip Cornford[i], Olalekan Lee Aiyegbusi[l], Mieke Van Hemelrijck[d], James N'Dow[e], Alberto Briganti[c] *and PIONEER Consortium*

[a]Department of Medicine and Surgery, Vita Salute San Raffaele University, Milan, Italy
[b]Department of Clinical and Health Psychology, IRCCS San Raffaele Hospital, Milan, Italy
[c]Unit of Urology/Division of Oncology, URI, IRCCS San Raffaele Hospital, Milan, Italy
[d]Translational and Oncology Research (TOUR), King's College London, Faculty of Life Sciences and Medicine, London, UK
[e]Academic Urology Unit, University of Aberdeen, Aberdeen, UK
[f]MRC North West Hub for Trials Methodology Research, University of Liverpool, a member of Liverpool Health Partners, Liverpool, UK
[g]Global Epidemiology, Bayer HealthCare Pharmaceuticals Inc., Whippany, USA
[h]Department of Urology, University Hospital, St. Etienne, France
[i]Liverpool University Hospitals NHS Trust, Liverpool, UK
[l]Centre for Patient-Reported Outcomes Research, College of Medical and Dental Sciences, University of Birmingham, Birmingham, UK
**both authors equally contributed*

Word count (abstract): 305**; Word count (text):** 4358; **Text pages**: 18; **Tables:** 3; **Figures**: 1; **Supplementary tables**: 10; **Appendix files:** 3
Running head: PROMs in localized Prostate Cancer
Keywords: Prostatic Neoplasms; COSMIN; PROMs; Quality of life; Erectile dysfunction; localized; prostate cancer

Corresponding author:
Giorgio Gandaglia, MD
Division of Oncology/Unit of Urology
Urological Research Institute
IRCCS Ospedale San Raffaele; Milan, Italy
Vita-Salute San Raffaele University
Via Olgettina, 60 – 20132 Milan, Italy
Tel: +39 02.2643.4066; Fax: +39 02.2643.7298
E-mail: gandaglia.giorgio@hsr.it

ABSTRACT

Context: Prostate cancer (PCa) is the second most common cancer in men worldwide. Urinary, bowel, and sexual function, as well as hormonal symptoms and Health-Related Quality of Life (HRQoL) were prioritised by patients and professionals as part of a core outcome set for localised PCa regardless of the treatment type.

Objective: To systematically review the measurement properties of patient reported outcome measures (PROMs) used in localised PCa and recommend PROMs for use in routine practice and research settings.

Evidence acquisition: The psychometric properties of PROMs measuring functional and HRQoL domains used in Randomized Controlled Trials (RCTs) including patients with localised PCa were assessed according to the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) methodology. Medline and Embase was searched to identify publications that evaluated psychometric properties of the PROMs. Characteristics and methodological quality of the included studies were extracted, tabulated and assessed according to the COSMIN criteria.

Evidence synthesis: Overall, 27 studies evaluating the psychometric properties of the EPIC, UCLA-PCI, EORTC QLQ-C30, EORTC QLQ-PR25, IIEF, SF-36 and SF-12 were identified and included in the systematic review. The EPIC and EORTC QLQ C-30, a general module which assesses patients' physical, psychological, and social functions, were characterized by high internal consistency (Cronbach's α=0.46-0-96 and 0.68-0.94 respectively) but low content validity. The EORTC QLQ-PR25, which is primarily designed to assess PCa-specific health-related quality of life, had a moderate content validity and internal consistency (Cronbach's α=0.39-0-87). The UCLA-PCI was characterized by moderate content validity and high internal consistency (Cronbach's α=0.21-0-94). However, it does not directly assess hormonal symptoms, where the EORTC QLQ-PR25 does.

Conclusion: The tools with the best evidence for psychometric properties and feasibility for use in routine practice and research settings to assess PROMs in patients with localized PCa were the EORTC QLQ-C30 and QLQ-PR25. Since the EORTC QLQ-C30 is a general module that does not directly assess PCa-specific issues, it should be adopted in conjunction with the EORTC PR-25 module.

Patient summary: In this paper, we reviewed and appraised the measurement properties of PROMs used with localised PCa. We found good evidence to suggest the use of the EORTC QLQ-C30 and QLQ-PR25 to measure core domains and HRQoL.

INTRODUCTION

Prostate Cancer (PCa) is the second most common solid cancer and the fifth leading cause of cancer death in men worldwide [1]. This disease is characterized by a relatively long natural history, where a substantial proportion of PCa patients die from causes other than the disease itself [2]. Therefore, the assessment and monitoring of treatment-related side effects as well as Health-Related Quality of Life (HRQoL) play a major role in the management of PCa patients. Although metastases-free and overall survival represent the main outcomes when assessing the efficacy of therapeutic approaches, there is increasing awareness of the importance of measuring what were previously regarded as softer outcomes (namely, side effects and HRQoL using Patient-Reported Outcome Measures [PROMs]) [3]. This is crucial when considering that cancer patients have the possibility of trading HRQoL for length of life [4].

Several PROMs have been proposed and are currently used for assessing different domains (i.e., urinary, sexual, and bowel function) and quality of life in localised PCa. However, the validity of these tools has been poorly addressed so far and limited data are available on the appropriate PROMs that should be adopted in prospective studies and randomized controlled trials (RCTs). Moreover, a systematic comprehensive assessment of the psychometric properties of PROMs in patients with localized PCa is still missing. PIONEER is a European network of excellence for big data in PCa and is part of the Innovative Medicine Initiative's (IMI's) "Big Data for Better Outcomes" programme [5]. The overall aim of PIONEER is to improve PCa care across Europe through the application of big data analytics [5]. One of the many PIONEER objectives was the further development of the Core Outcome Set (COS) for localised PCa [6]. Consensus on recommendations for the definitions and measures of the clinician reported outcomes has been completed by the PIONEER consortium. The patient reported outcomes, the focus of this paper, are: overall QoL; hormonal symptoms; sexual function; urinary function; bowel function; stress urinary incontinence; and faecal incontinence as previously defined [6].

The aim of this systematic review is to critically appraise, compare, and summarize the measurement properties of PROMs used in men with localised PCa following the COnsensus-based Standards for the selection of health Measurement Instruments (COSMIN) guidelines and to summarise the feasibility aspects associated with the tools (e.g., time to complete, fee payable, training required). Our findings should help physicians and researchers in selecting the most appropriate PROMs to use for HRQoL assessment in patients with localised PCa.

EVIDENCE ACQUISITION

The systematic review was conducted according to a pre-defined protocol based on the COSMIN methodology [7,8]. The review is reported in accordance with the PRISMA statements [9]. Several systematic reviews of HRQoL instruments in cancer patients adopted the COSMIN Checklist to evaluate PROMs [10,11]. For this systematic review, instruments/measures to record HRQoL will be referred to as PROMs, which are typically defined as questionnaires used to assess every construct considered. The COSMIN approach defines the measurement properties important to PROMs: validity, responsiveness, and reliability. To evaluate these psychometric properties, the study was divided into four steps: identification, prioritization, assessment, and summary of all PROMs considered. Four of the members of our research team (A.D., F.F., E.S.S., S.M.F.) undertook all stages of the process following the original electronic search, working independently of each other, and then comparing outcomes at each stage of the process. Any disagreements were resolved by discussion between the members of the research team.

Search Strategy

First, to identify a comprehensive list of PROMS used in localised PCa intervention effectiveness trials, we updated the systematic review by MacLennan et al. and identified the PROMs used in RCTs focused on non-metastatic PCa patients [12]. We used the same inclusion/exclusion criteria as a pragmatic decision. We aimed at selecting studies which are likely to influence clinical practice and we respected the hierarchy of evidence of Oxford Centre for Evidence-Based Medicine guidelines [32] focusing on RCTs. We extracted data on the core functional and QoL outcomes and domains: urinary function; bowel function; sexual function; overall quality of life. Studies including a mixed population of patients with localised or non-metastatic locally advanced PCa were included. The PROMS identified are listed in Supplementary Table 1.

After identifying the eligible PROMS, a systematic review was performed to identify publications developing or evaluating psychometric properties of the identified PROMs in patients with localised PCa. The search evaluated studies published up to November 2019 without other time restrictions. Databases searched included MEDLINE (PubMed) and EMBASE using the PROM filter developed for PubMed by the Patient Reported Outcomes Measurement Group, University of Oxford and a highly sensitive validated search filter for finding studies on measurement properties available on the COSMIN website [13,14]. Terms

were agreed by the research team and limited to English-language articles that had developed, validated or assessed the psychometric properties of the different PROMs [15].

Eligibility criteria

Inclusion criteria: (i) studies that have assessed the measurement properties of a PROM (development and validation papers); (ii) studies including localised PCa patients (mixed populations were acceptable so long as localised PCa patients were reported separately), studies including participants with other disease types were considered if the results for localised PCa participants were reported separately; (iii) adequate evidence of content validity such as a qualitative study of the construct of interest in the target population.

Exclusion criteria: (i) studies assessing non-health related PROMs (e.g., treatment satisfaction); (ii) studies where the PROM was used exclusively to evaluate Health Related Quality of Life in PCa patients and not reporting on psychometric properties; (iii) studies including mainly metastatic PCa patients only or mixed populations with no separate reporting of localised patients; (iv) meeting abstracts, conference abstracts, editorials, and commentaries; (v) non-English language.

To identify PROMs suitable for the assessment of psychometric properties, the team used search masks with specific key words that can be seen in Appendix 1.

Screening

Abstracts and full texts were screened against the eligibility criteria by two reviewers independently between November 2019 and July 2020. Any disagreements were resolved by a third reviewer. Full-text papers were further explored independently by each of the four other reviewers to identify the final articles for inclusion in the review. The results from the database search and the study selection process are presented in Table 1.

Data Extraction

Data were extracted for the following characteristics from each study: PROM(s) used, construct(s) measured, number of (sub)scales, number of items, description and version of the PROM, recall period, scoring information, time required to complete the PROM, information on population, training required for the administration of the PROM, mode of administration, order form, type of license, number of published studies using the instrument, highest COSMIN rating, additional comments and online example. This information was included in summary cards for each PROM (Supplementary Tables 2, 3, 4, 5, 6, 7, 8). The

summary card format was based on Turnbull et all's PROM assessment research in the Critical Care setting (https://www.improvelto.com/instruments/) [16].

Appraising Methodological Quality

To evaluate the methodological quality of studies, the COSMIN 4-point Checklist was used. The checklist criteria cover nine measurement properties, and the included studies' measurement properties were assessed using COSMIN Risk of Bias Checklist. There are three domains containing various measurement properties: reliability ("internal consistency", "reliability", and "measurement error"), validity ("content validity", "structural validity", "hypothesis testing", "cross-cultural validity", and "criterion validity"), and "responsiveness" [15]. Appendix 2 shows the measurement properties and definitions [17].

Eligible studies were rated as "very good", "adequate", "doubtful", "inadequate" and "not applicable" for each measurement property [8]. The overall rating of the quality of each study on a measurement property was determined by the lowest rating of any standard: "the worst score counts" [16]. All results per study measurement property of a PROM were quantitatively pooled or summarized against the criteria for good measurement properties to get the overall ratings.

Reporting of Psychometric Results

All PROMs were graded on the quality of the evidence. The internal consistency depends on the available evidence for structural validity because the prerequisite for the interpretation of internal consistency is the unidimensionality. The quality of evidence for structural validity was the starting point for determining the quality of evidence for internal consistency. However, Cronbach's alpha is difficult to interpret since it is not based on a unidimensional scale [18]. As such, the COSMIN manual for systematic reviews of PROMs recommends ignoring results of studies on internal consistency of scales which are not unidimensional. The most important psychometric property is the "content validity", which reflects if the items of the PROM are "relevant", "comprehensive", and "comprehensible" to the construct of interest and study population. This is why we considered this a vital property and used it as a threshold for inclusion (i.e., if a PROM could not demonstrate adequate content validity for the domain of interest [e.g. urinary function] in our target population it was not considered further). To evaluate the content validity a subjective judgment by the reviewers is required on the adequacy of the initial qualitative work in the target population to identify the constructs of importance. This should include the PROM development study, the quality and

results of additional content validity studies on the PROMs (where available), and a subjective rating of the content of the PROMs [19].

Levels of Evidence Appraisal

To determine the overall quality of each measurement property, a levels-of-evidence appraisal was undertaken. This process produced a final rating for each PROM for each measurement property. By using the Excel template provided by the COSMIN group, it was possible to evaluate the quality of each PROM's development and to standardize the evaluation of the quality of the PROM's design. Following this template, a rating system from "very good" to "inadequate" was used. COSMIN Checklist criteria were used to evaluate the quality of content validity studies, concerning "comprehensiveness", "comprehensibility", and "relevance" of each PROM item. To evaluate the content validity, the results of the single studies on PROMs development were rated. PROMs were then categorized into three categories to allow for an evidence-based recommendation: "A" rated are PROMs with evidence for sufficient content validity and at least low-quality evidence for sufficient internal consistency. PROMs rated "A" can be recommended for use and results can be trusted; "B" are PROMs that have potential to be recommended but they require further research; "C" are PROMs with high quality evidence for an insufficient measurement property.

All available studies were qualitatively summarized to determine whether overall, the "relevance", "comprehensiveness", "comprehensibility", and overall content validity of the PROMs are "sufficient", "insufficient", or "indeterminate". The overall ratings were accompanied by a grading for the quality of the evidence. "High" (+), or "moderate" (±), or "low" (?), or "very low" (-) are the levels of evidence rating. Levels of evidence criteria are presented in Appendix 3.

EVIDENCE SYNTHESIS

The most frequently used PROMs in RCTs of men with localised PCa were the IIEF (n=50), the EPIC (n=26) and the ICIQ (n=11). Supplementary Table 1 depicts the list of the PROMs identified in the systematic review as well as the corresponding publications. Some of the instruments used for the assessment of quality of life in PCa patients are not present among the PROMs suitable for psychometric evaluation. The reason for this resides in the rigorous methodology applied in our systematic review, where PROMs were considered exclusively if they were adopted in RCTs focusing on patients with localized PCa. In addition, some studies evaluating additional PROMs were excluded since they did not pass the inclusion

criteria in the full text screening phase. The characteristics of included studies are presented in Supplementary Table 2. After removal of duplicates and abstract screening, a total of 113 full texts were selected for further examination. Among those, 27 met the inclusion criteria and were evaluated according to the COSMIN Checklist (Figure 1). At the end of the screening phase we identified two generic PROMs: 36-item Short-Form health survey (SF-36), 12-item Short-Form health survey (SF-12); one cancer-generic PROM: European Organization for Research and Treatment of Cancer quality-of-life, 30 item core questionnaire (EORTC QLQ-C30) [20], three Prostate Cancer-specific PROMs: Expanded Prostate Cancer Index Composite (EPIC) [21,22]; EORTC 25 item core questionnaire (QLQ-PR25); UCLA Prostate Cancer Index (UCLA-PCI) [23]; and one specific PROM for erectile function: International Index of Erectile Function (IIEF) [24]. The PCa-specific PROM (EPIC) was the most frequently evaluated in validation studies (n=8) followed by the EORTC QLQ-C30 (n=7) and the UCLA-PCI (n=3). A summary of the number of items and concepts assessed in each of the evaluated PROMs is presented in Supplementary table 2.

Table 2 presents COSMIN Checklist scores, which assess the methodological quality of studies that reported COSMIN measurement properties for each PROM. Internal consistency was the most frequently reported property (21 studies). The reliability was calculated in 17 studies, the structural validity and the hypothesis testing in 9, and the responsiveness and the cross-cultural validity in 7. While the least frequently reported property was measurement error (2 studies). The content validity was assessed for each study by the reviewers following the COSMIN criteria. The best performing properties were the internal consistency and the hypothesis testing with 8 of the 21 and 5 of the 9 studies receiving a score of “very high”, respectively. In addition, 4 of the 17 studies had a “very high” reliability, and 1 of the 7 studies had a “very high" cross-cultural validity. The properties with the worst performance were the content validity and the structural validity, where 6 of the 25 and 3 of 9 studies scored “very low”, respectively. The internal consistency was presented as Cronbach’s α, while the reliability was calculated using Pearson's or Spearman’s correlation or, less frequently, using the Interclass Correlation (ICC). Structural validity was assessed by either exploratory or confirmatory factor analysis.

Following the COSMIN criteria, the PROMs were then placed into three categories to recommend the most suitable PROMs for use in clinical practice and research settings. We categorized as “A” (i.e., can be recommended for use and results can be trusted) the UCLA-PCI and the EORTC QLQ-PR25, as “B” (i.e., have potential to be recommended but they require further research) the EPIC and the EORTC QLQ-C30, and as “C” (i.e., high quality

evidence for an insufficient measurement property) the SF-12, the IIEF, and the SF-36. We scored the summarized content validity for each PROM, averaging across all the content validity scores previously assigned to each individual study.

Based on the COSMIN methodology, greater importance is given to psychometric properties such as construct validity and internal consistency scores. Nevertheless, the other parameters remain equally important and desirable with all the instruments evaluated. Table 3 shows the summarized measurement properties for each PROM. The UCLA-PCI showed moderate content validity and high internal consistency (summarized Cronbach's α=0.21-0.94); the EORTC QLQ-PR25 had moderate content validity and moderate internal consistency (summarized Cronbach's α=0.39-0.87). Both the EPIC and the EORTC QLQ-C30 show low content validity and high internal consistency (summarized Cronbach's α=0.46-0.96, and 0.68-0.94, respectively). Finally, the SF-12, IIEF and SF-36 had low content validity while the internal consistency was not calculated in the studies of the SF-12 and IIEF and was scored as moderate the study of the SF-36 (summarized Cronbach's α=0.77-0.93). An overview of each PROM along with the feasibility aspects such as time to complete, licensing requirements and administrator training [33], is displayed in Supplementary Table 4-10. The concept 'feasibility' is related to the concept of 'clinical utility', whereas feasibility focuses on PROMs, clinical utility refers to an intervention [34]. Interpretability and feasibility are not measurement properties because they do not refer to the quality of a PROM. However, they are considered important aspects for a well-considered selection of a PROM. As can be seen in Supplementary Table 3, most of the questions have free access, so they are easily available apart from EORTC-30 and EORTC QLQ-PR25 which need a user agreement to get permission to use them. They all take a few minutes to fill in and no specific training is needed before they are adopted. The UCLA-PCI is categorised as 'A', which means that in addition to the statistically significant features, it has good feasibility (i.e., it takes little time, requires no specific training for administration, does not require access costs and it is a specific tool to measure quality of life in the setting of PCa). However, the UCLA-PCI does not directly assess hormonal symptoms, where the EORTC QLQ-PR25 does, therefore, the EORTC PR-25 module is the most appropriate for use in the localised PCa population together with the EORTC QLQ-C30 to assess generic cancer patients' HRQoL.

DISCUSSION

Our study rigorously applied the COSMIN methodology to assess the psychometric properties of PROMs to be used in research studies and, ideally, in clinical practice for

localised PCa patients in the context of a wider COS development project that included patients, healthcare professionals and industry partners. The psychometric properties of all selected PROMs were assessed using a rigorous and standardized tool after performing a systematic review of the literature. Overall, 27 studies that reported psychometric properties of PROMs met the inclusion criteria. Although none of them met all COSMIN standards for methodological quality, the “best fit” method was adopted to make our assessment objective. This represents the first attempt to systematically assess PROMs to be used in clinically localized PCa using a validated rigorous approach.

The results of our study are several-fold. First, although our systematic review identified twelve PROMs used in RCTs on localised PCa, studies assessing the psychometric properties were not available for five of them. For example, the ICIQ and IPSS are commonly used in localised PCa and were adopted in 11 and 7 RCTs, respectively. However, no data is available regarding their psychometric properties. Although both these questionnaires are considered as reliable PCa-specific tools by clinicians and typically used in retrospective studies and prospective trials, our observations highlight that evidence-based recommendations are urgently needed to support the PROM that clinicians and researchers should use to evaluate patients with localized disease. This will have a relevant impact to assess the external validity of ongoing trials, to compare results among different studies evaluating patient-reported outcomes, and to directly address the challenges of outcome reporting heterogeneity. Finally, our observations highlight the need for more research activities aimed at validating the psychometric properties of available PROMs assessing cancer-specific domains such as urinary incontinence and erectile function.

Second, seven out of the eleven identified PROMs had only limited data available on their psychometric properties. Among those, the Short Form-36 Health Survey (SF-36) and the Short Form-12 Health Survey (SF-12) provide information about patients’ HRQoL in general, but not specifically about the core domains of urinary, bowel and sexual function. The SF-36, the SF-12, and the IIEF are accepted as being valid, reliable, and sensitive in a wide range of health problems [24,25]. The IIEF covers sexual function but not urinary or bowel function. However, the SF-12, the SF-36, and the IIEF may lack sensitivity in measuring PCa specific issues. They show a high-quality evidence of an insufficient measurement property and we cannot recommend these tools for use in routine practice or research. On the other hand, the EORTC QLQ-C30 and the EPIC have been scored as “good” tools for assessing the quality of life in PCa patients, but they showed a low content validity and a high internal consistency. Therefore, they have potential to be recommended but they

require further research. Of note, the adoption of the EPIC has been recommended by International Consortium for Health Outcomes Measurements (ICHOM). However, further studies are needed before its use can be routinely recommended in the ideal scenario. The scores of the EORTC QLQ-C30 are an adequate reflection of the dimensionality of the construct to be measured and an adequate interrelatedness among the items. Moreover, the performance of the items on translated or culturally adapted EORTC QLQ-C30 are an adequate reflection of the performance of the items of the original version of the PROM and the total variance in the measurements of this PROM is due to 'true' differences between patients. Even though it scored 'B' overall, the EORTC QLQ-C30 had the highest quality of evidence available for assessing HRQoL in localised PCa patients and therefore we recommend it for assessing this domain.

Both the EORTC QLQ-PR25 and the UCLA-PCI performed well and have the most positive COSMIN ratings. Therefore, we classified them in group "A" since the first one shows a moderate content validity and a high internal consistency, and the second one shows a moderate content validity and a moderate internal consistency [26]. The EORTC QLQ-PR25 reflects mostly the performance of the items for PCa quality of life. The UCLA-PCI might have good internal structure (structural validity, internal consistency, cross-cultural validity) because it refers to how the different items in the PROM are related, which is important to help understand how items might be combined into a scale or subscale. This PROM reflects specifically a good compatibility for evaluating the HRQoL in PCa patients in routine clinical practice. However, the UCLA-PCI does not directly assess hormonal symptoms, where the EORTC QLQ-PR25 does, therefore, considering hormonal symptoms is one of the core outcomes we recommend, the EORTC PR-25 module is the most appropriate for use in the localised PCa population [6]. Of note, this module should be used in conjunction with the general module EORTC QLQ-C30 which assesses patients' physical, psychological, and social functions.

As highlighted in the summary cards, beyond the psychometric properties there are feasibility aspects that should be considered. These relate, for example, to the availability of the PROMs and their costs, as well resources required, such as staff training and time to complete the tool. These resources may differ between the research and routine practice settings. Considering the psychometric and feasibility information together, the PIONEER consortium recommends that the EORTC QLQ-30 and the EORTC-PR25 should be used in both research and routine care settings. Previous studies recommended different PROMs to be implemented in the clinical practice (i.e., EPIC, PORPUS and PC-QoL [30]. However,

the evidence beyond these recommendations is based on historic cohorts and considered only PCa patients with early-stage disease. Conversely, our systematic review considered contemporary instruments developed up to 2014 and takes advantage from a standardized systematic approach using the COSMIN methodology to identify the strongest questionnaires at the psychometric level.

Taken together, the recommended PROMs provide the most coverage of the HRQoL and functional domains of the COS whilst being the soundest measures available of these core outcome domains in localised PCa. These tools should be administered at different time-points at baseline and during follow-up according to the type of treatment delivered as recommended by ICHOM. Researchers and clinicians should be encouraged to use the EORTC QLQ-30 and EORTC-PR25. Using the same tools to measure the main domains of results ensures less heterogeneity in the evidence base. The use of common outcome measurement instruments will facilitate the interpretation of the evidence across different RCTs and will facilitate to summarize evidence meta-analyses with a consequent impact on recommendation-making in clinical guidelines. Furthermore, the use of the same measures in routine clinical practice and other real-world evidence data sources will facilitate the integration and analysis in Big Data platforms such as PIONEER. This has powerful implications for benchmarking via audit and feedback from both 'value-driven' [27] and quality improvement approaches [28]. Beside disease-specific validated tools, generic measures are useful for comparability, for their greater practicality and may be relevant to the multiple stakeholders involved in PCa management. For example, policymakers may look to generic instruments (and QALYs) to ascertain whether to fund one therapy, not just against other potential PCa-interventions, but compared to interventions in other settings to ensure that cost-effective care is delivered to maximize health outcomes. Under this light, further research is needed in this setting to validate generic PROMs in localized PCa patients which should be considered in addition to disease specific tools.

Although our initial protocol also included a Delphi survey after the assessment of the PROMs using the COSMIN criteria in localized PCa, the identification of only two PROMs (namely, EORTC QLQ-PR25 and UCLA-PCI) graded as "A" limited the potential value of a formal Delphi survey. Indeed, it is likely that physicians would indicate as appropriate the use of most adopted PROMS rather than the ones with a more solid construct and validity. On the other hand, our work should inform physicians regarding the availability of validated PROMs assessed via a systematic approach in the setting of localized PCa regardless of their preferences. The mere recommendation to use these validated tools would not

necessarily translate into their adoption by clinicians in the everyday practice and, therefore, implementation measures might be necessary. Whilst the implementation of PROMs into standard care is still in development [29], useful guidance is already available for the reporting, analyses, and translations of PROMs. For instance, the EORTC has developed a manual for the use of their measures in daily clinical practice [31].

Some potential limitations of our study should be discussed. Firstly, our systematic review was restricted to the English language, which may have introduced a language bias. At the same time, our aim was to compare the psychometric aspects of HRQoL instruments, and there is no evidence to suggest that this objective was influenced by the language restriction. One of the main strengths of this study is that we have been able to categorize the PROMs in the "A", "B", and "C" categories according to the different psychometric properties assessed. In addition, the use of the COSMIN checklist is a strength as it allows the content validity of PROMs to be calculated in a standardised manner through quality scores of measurement properties based on systematic reviews. This should guide physicians in the decision regarding which PROMs should be adopted in their clinical practice regardless of their subjective preferences.

CONCLUSIONS

The psychometric properties of several PROMs which are currently used for patients with localised PCa have been poorly assessed so far. Although the UCLA-PCI was characterized by moderate content validity and high internal consistency, it does not directly assess hormonal symptoms while the EORTC QLQ-PR25 does. Since the EORTC QLQ-C30 is a general module that does not directly assess PCa-specific issues, it should be adopted in conjunction with the EORTC PR-25. Therefore, the PIONEER consortium recommends that the EORTC QLQ-30 and the EORTC-PR25 should be used in both research and routine care settings to measure the core domains of urinary, bowel, and sexual function, hormonal symptoms and HRQoL.

REFERENCES

[1] Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer Statistics, 2021. CA: A Cancer Journal for Clinicians 2021; 71:7–33. https://doi.org/10.3322/caac.21654.

[2] Briganti A, Karnes RJ, Gandaglia G, Spahn M, Gontero P, Tosco L, et al. Natural history of surgically treated high-risk prostate cancer. Urologic Oncology: Seminars and Original Investigations 2015;33. https://doi.org/10.1016/j.urolonc.2014.11.018.

[3] Black N. Patient reported outcome measures could help transform healthcare. BMJ (Online) 2013;346. https://doi.org/10.1136/bmj.f167.

[4] Shrestha A, Martin C, Burton M, Walters S, Collins K, Wyld L. Quality of life versus length of life considerations in cancer patients: A systematic literature review. Psycho-Oncology 2019;28:1367–80. https://doi.org/10.1002/pon.5054.

[5] Omar MI, Roobol MJ, Ribal MJ, Abbott T, Agapow PM, Araujo S, et al. Introducing PIONEER: a project to harness big data in prostate cancer research. Nature Reviews Urology 2020;17:351–61. https://doi.org/10.1038/s41585-020-0324-x.

[6] MacLennan S, Williamson PR, Bekema H, Campbell M, Ramsay C, N'Dow J, et al. A core outcome set for localised prostate cancer effectiveness trials. BJU International 2017;120:E64–79. https://doi.org/10.1111/bju.13854.

[7] Mokkink LB, de Vet HCW, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, Terwee CB. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. Qual Life Res. 2018 ; 27: 1171-1179

[8] Terwee CB, Mokkink LB, Knol DL, Ostelo RWJG, Bouter LM, de Vet HCW. Rating the methodological quality in systematic reviews of studies on measurement properties: A scoring system for the COSMIN checklist. Quality of Life Research 2012;21:651–7. https://doi.org/10.1007/s11136-011-9960-1.

[9] Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (prisma-p) 2015: Elaboration and explanation. BMJ (Online) 2015;349. https://doi.org/10.1136/bmj.g7647.

[10] Tax C, Steenbergen ME, Zusterzeel PLM, Bekkers RLM, Rovers MM. Measuring health-related quality of life in cervical cancer patients: A systematic review of the most used questionnaires and their validity. BMC Medical Research Methodology 2017;17. https://doi.org/10.1186/s12874-016-0289-x.

[11] Wong CKH, Chen J, Yu CLY, Sham M, Lam CLK. Systematic review recommends the European Organization for Research and Treatment of Cancer colorectal cancer-specific module for measuring quality of life in colorectal cancer patients. Journal of Clinical Epidemiology 2015;68:266–78. https://doi.org/10.1016/j.jclinepi.2014.09.021.

[12] MacLennan S, Bekema HJ, Williamson PR, Campbell MK, Stewart F, MacLennan SJ, et al. A core outcome set for localised prostate cancer effectiveness trials: Protocol for a systematic review of the literature and stakeholder involvement through interviews and a Delphi survey. Trials 2015;16. https://doi.org/10.1186/s13063-015-0598-0.

[13] Prinsen CAC, Vohra S, Rose MR, Boers M, Tugwell P, Clarke M, et al. How to select outcome measurement instruments for outcomes included in a “Core Outcome Set” - a practical guideline. Trials 2016;17. https://doi.org/10.1186/s13063-016-1555-2.

[14] Aiyegbusi OL, Kyte D, Cockwell P, Marshall T, Keeley T, Gheorghe A, et al. Measurement properties of patient-reported outcome measures (PROMs) used in adult patients with chronic kidney disease: A systematic review protocol. BMJ Open 2016;6. https://doi.org/10.1136/bmjopen-2016-012014.

[15] Mokkink LB, de Vet HCW, Prinsen CAC, et al. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. Qual Life Res. 2018 ; 27: 1171-1179

[16] Needham DM, Sepulveda KA, Dinglas VD, Chessare CM, Friedman LA, Bingham CO, et al. Core outcome measures for clinical research in acute respiratory failure survivors an international modified delphi consensus study. American Journal of Respiratory and Critical Care Medicine 2017;196:1122–30. https://doi.org/10.1164/rccm.201702-0372OC.

[17] Terwee CB, Prinsen CAC, de Vet HCW, Bouter LM, Alonso J, Westerman MJ, et al. COSMIN methodology for assessing the content validity of Patient-Reported Outcome Measures (PROMs). User manual, version 1. User Manual, Version 1 2017.

[18] Terwee CB, Prinsen CAC, Chiarotto A, Westerman MJ, Patrick DL, Alonso J, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. Quality of Life Research 2018;27:1159–70. https://doi.org/10.1007/s11136-018-1829-0.

[19] Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. Quality of Life Research 2018;27:1147–57. https://doi.org/10.1007/s11136-018-1798-3.

[20] Arraras JI, Wright S, Greimel E, Holzner B, Kuljanic-Vlasic K, Velikova G, et al. Development of a questionnaire to evaluate the information needs of cancer patients:

The EORTC questionnaire. Patient Education and Counseling 2004;54:235–41. https://doi.org/10.1016/S0738-3991(03)00240-4.

[21] Wei JT, Dunn RL, Litwin MS, Sandler HM, Sanda MG. Development and validation of the Expanded Prostate Cancer Index Composite (EPIC) for comprehensive assessment of health-related quality of life in men with prostate cancer. Urology 2000;56:899–905. https://doi.org/10.1016/S0090-4295(00)00858-X.

[22] Szymanski KM, Wei JT, Dunn RL, Sanda MG. Development and validation of an abbreviated version of the expanded prostate cancer index composite instrument for measuring health-related quality of life among prostate cancer survivors. Urology 2010;76:1245–50. https://doi.org/10.1016/j.urology.2010.01.027.

[23] Litwin MS, Hays RD, Fink A, Ganz PA, Leake B, Brook RH. The UCLA Prostate Cancer Index: Development, Reliability, and Validity of a Health-Related Quality of Life Measure. Medical Care 1998;36:1002–12. https://doi.org/10.1097/00005650-199807000-00007.

[24] Rosen RC, Riley A, Wagner G, Osterloh IH, Kirkpatrick J, Mishra A. The international index of erectile function (IIEF): A multidimensional scale for assessment of erectile dysfunction. Urology 1997;49:822–30. https://doi.org/10.1016/S0090-4295(97)00238-0.

[25] Cappelleri JC, Rosen RC, Smith MD, Mishra A, Osterloh IH. Diagnostic evaluation of the erectile function domain of the International Index of Erectile Function. Urology 1999;54:346–51. https://doi.org/10.1016/S0090-4295(99)00099-0.

[26] Matsuda T, Marche H, Grosclaude P, Clement S. Participation behavior of bladder cancer survivors in a medical follow-up survey on quality of life in France. European Journal of Epidemiology 2004;19:313–21. https://doi.org/10.1023/B:EJEP.0000024703.54660.e5.

[27] Martin NE, Massey L, Stowell C, Bangma C, Briganti A, Bill-Axelson A, et al. Defining a standard set of patient-centered outcomes for men with localized prostate cancer. European Urology 2015;67:460–7. https://doi.org/10.1016/j.eururo.2014.08.075.

[28] Improving healthcare quality in Europe: Characteristics, effectiveness and implementation of different strategies [Internet] - PubMed n.d. https://pubmed.ncbi.nlm.nih.gov/31721544/ (accessed April 29, 2021).

[29] Montgomery N, Howell D, Ismail Z, Bartlett SJ, Brundage M, Bryant-Lukosius D, et al. Selecting, implementing and evaluating patient-reported outcome measures for

routine clinical use in cancer: the Cancer Care Ontario approach. Journal of Patient-Reported Outcomes 2020;4. https://doi.org/10.1186/s41687-020-00270-1.

[30] Schmidt, S., Garin, O., Pardo, Y., Valderas, J. M., Alonso, J., Rebollo, P., Rajmil, L., Garcia-Forero, C., Ferrer, M., & EMPRO Group (2014). Assessing quality of life in patients with prostate cancer: a systematic and standardized comparison of available instruments. Quality of life research: an international journal of quality of life aspects of treatment, care and rehabilitation, 23(8), 2169–2181. https://doi.org/10.1007/s11136-014-0678-8

[31] https://www.eortc.org/app/uploads/sites/2/2018/02/EORTC_QLQ_Clinical_Practice_User_Manual-1.0.pdf n.d.

[32] Oxford Centre for Evidence-based Medicine - Levels of Evidence (March 2009). Centre for Evidence-Based Medicine. 2009. http://www.cebm.net/ index.aspx?o=1025. Accessed 16 Feb 2015

[33] Boers M, Kirwan JR, Tugwell P, Beaton D, Bingham CO, III, Conaghan PG, et al. The OMERACT handbook: OMERACT; 2015

[34] Smart A. A multi-dimensional model of clinical utility. International journal for quality in health care : journal of the International Society for Quality in Health Care. 2006;18(5):377-82

Funding

PIONEER is funded through the IMI2 Joint Undertaking and is listed under grant agreement No. 777492. This joint undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA.

OLA declares funding from the National Institute for Health Research (NIHR) Birmingham Biomedical Research Centre (BRC), the NIHR Applied Research Centre (ARC), at the University of Birmingham and University Hospitals Birmingham NHS Foundation, Innovate UK (part of UK Research and Innovation), Gilead Sciences Ltd, and Janssen Pharmaceuticals, Inc. OLA declares personal fees from Gilead Sciences Ltd, GlaxoSmithKline (GSK) and Merck outside the submitted work.

PW is a National Institute for Health Research (NIHR) Senior Investigator (award number NF-SI_0513-10025 supporting this work). The views expressed in this article are those of the author(s) and not necessarily those of the NIHR, or the Department of Health and Social care.

Competing interests

None

Acknowledgements

We would like to thank Caroline Terwee and Sanna Prinsen from the COSMIN initiative for their advice on our study design and application of the COSMIN tools.