患者报告结局在药物临床研究中应用的指导原则

（试行）

December 2021

目 录

患者报告结局在药物临床研究中应用的指导原则（试行）

I. Introduction

Clinical outcomes are the core basis for evaluating the benefits and risks of drug treatment. How Accurate, reliable, and complete observation of clinical outcomes is crucial. patient reported outcomes (patient-reported outcome, PRO) is one of the forms of clinical outcomes. It is increasingly used in drug registration clinical research. In addition, as patients patient-focused drug development (PFDD) With the continuous development of concepts and practices, the patient's body is captured throughout the drug life cycle. experience, insights, needs and other data and effectively integrate them into drug development and evaluation is receiving increasing attention in price, clinical outcome (clinical outcome assessments (COA), particularly where patient-reported outcomes can reflect patient experience is an important component of patient-centered drug development.

This guidance is intended to clarify the definition of PROs and their use in drug registration studies. Scope of application, PRO measurement, especially general principles for scale development and use, PRO Quality control of data collection, matters needing attention in data analysis and interpretation, and communication with regulatory authorities, etc., to provide sponsors with reasonable information during drug registration studies. Use PRO data to provide guidance.

This guidance applies to the use of PROs as endpoints to support drug injection. Registered clinical research, including clinical trials and real-world studies.

2. Definition of patient-reported outcomes

Patient-reported outcomes are defined as any outcome directly reported by the patient and not

A person's modified or interpreted evaluation of their feelings about their disease and corresponding treatment.

PRO emphasizes patient self-reported outcomes. When patients do not have or lose their

Assessing competency may need to be completed by their guardian or a representative designated by the guardian

PROs are documented, but proxy bias should be fully assessed at this time.

The scale is a commonly used tool for PRO measurement and is mainly used for subjective measurement.

Such as pain, quality of life, etc., but existing scales cannot solve all subjective observations.

Quantity issues, such as certain symptoms (such as nausea) or clusters of symptoms. PRO data collection

There are two methods of paper records and electronic carriers. Use electronic means to record

PROs are called electronic patient-reported outcomes

outcomeÿePROÿÿ

3. Development, translation and improvement of patient-reported outcome measurement scales

In clinical research, once it is determined to use a scale to measure PRO, if there is no suitable

A scale that is suitable for research projects must be developed specifically for the research purpose; if it has

There are recognized Chinese scales suitable for research projects, which can be used directly after obtaining the copyright.

If there is a foreign language scale that is recognized as suitable for the research project, it needs to be developed

Use it after the official Chinese version is formed; if the existing scale is not completely suitable for research

When researching projects, it needs to be improved and used. How to choose a newer model among existing mature scales?

Suitable for the research project to be carried out, its scientific nature and operability need to be considered.

(1) Development of patient-reported outcome measurement scales

PRO measurement scales should be developed to reflect the patient's perspective, with emphasis on this

The clinical value of the scale, including the pertinence of efficacy evaluation and the interpretability of clinical significance

interpretation and guidance for treatment decisions. The development process of the scale is shown in Figure 2.

Scales are often developed for effectiveness evaluation, but may also target important safety

Events are developed, and the principles and processes are the same.

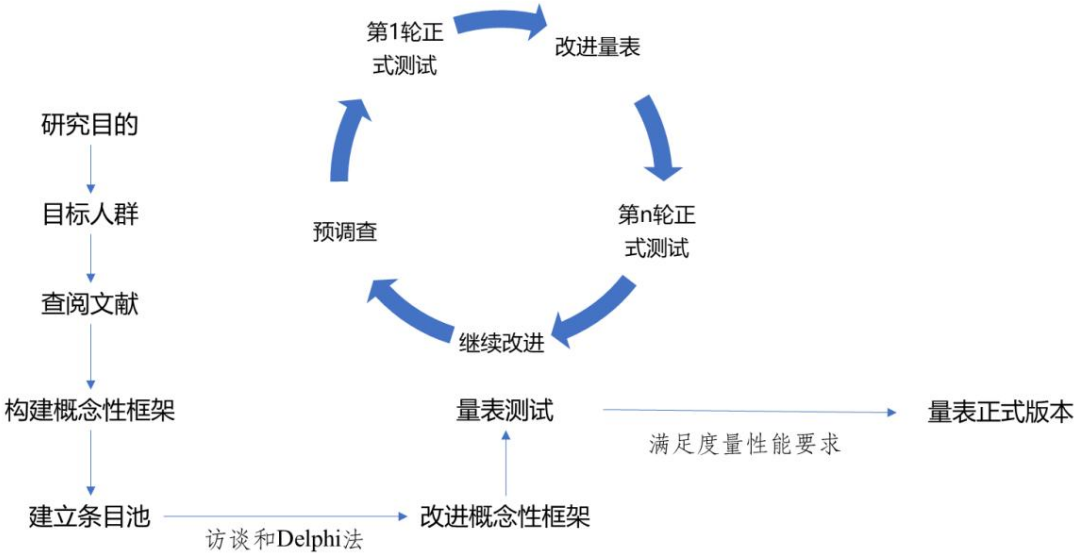

Figure 2 Schematic diagram of the scale development process

1. Build a conceptual framework

The structure of the scale includes primary structure, secondary structure and tertiary structure. Clinical research

In research, primary and secondary structures are more commonly used. Scales with first-level structure have single-item quantities

scales (e.g., Visual Analogue Pain Scale) and multi-item scales (e.g., Abbreviated Dry Mouth

dryness scale). The following takes the secondary structure scale as an example to illustrate.

The first level of the secondary structure scale is the dimension, and the second level is the item. Scale overview

The preliminary shaping of the conceptual framework is generally based on the developer's literature review, expert knowledge and

experience, patient interviews, and necessary research. The number and naming basis of the dimensions

The understanding setting of the research content, the number of items and the content of the items under each dimension are used to Reflect the connotation and importance of the dimension to which it belongs, for example, when each item is equally weighted, The number of entries under a dimension reflects the importance of the dimension.

2. Create an entry pool

The underlying structure of the scale is the items, which reflect the specific question content, and the dimensions It is conceptual. For subsequent item design, it is necessary to establish as rich as possible Entry pool, the source of the entry can be all possible ways, including literature, patients interviews with researchers and/or experts, scale development platforms in related fields, research and development reports, research Developer design, etc.

Item design is one of the core contents of scale development. If the entry pool is sufficient Rich and mature, most entries are generally obtained from the entry pool, but also Some items are designed by developers. Whenever possible, the problem statement should use Close-ended questions, avoid ambiguous words, double meanings, or tendencies Leading questions, double negative statements, and negative statements and patient reluctance questions to be answered; ceiling or floor effects of responses should also be avoided, as well as a Each entry asks two or more questions at the same time, etc. In terms of reading comprehension, try to use Commonly used expressions, the requirements for educational level should not be too high (such as having a primary school education) level of reading ability is sufficient).

3. Scaling method

The scale of items includes binary scale, grade scale (such as Likert scale), Continuous scaling (such as visual analog scale), graphic scaling and other methods, among which 5

Level Likert scaling method is most commonly used. How many levels of Likert scale should be used?

The measurement performance reaches the best standard.

4. Interview

After the developers have initially formed the conceptual framework of the scale, they first need to

Conduct patient interviews, expert interviews and/or expert surveys, and make adjustments based on expert feedback

overall conceptual framework. Patient interviews can help further ensure the quantity of patient-reported outcomes

The content validity of the form reflects the needs and opinions of patients. Experts investigate the main

The purpose is to verify the rationality of the structure, the accuracy of the item expressions, and the feasibility of the responses.

Sexuality and empowerment of dimensions and items. The weighting of dimensions and items is the most important aspect of scale development.

critical link. The expert survey method is usually implemented in more than one round to reach the expert

Until the opinions are relatively unified, especially the opinions on item empowerment.

5. Pre-investigation and formal investigation

The scale was developed after incorporating expert opinions to improve the initial conceptual framework.

Initial beta version, which then needs to be tested on the target population, and then based on the test

The trial results were improved and an official beta version was formed. Use the official beta version in your target audience

When conducting a survey, the sample size needs to be estimated based on the parameters of the pre-survey.

The improvements of the beta version are also adjusted based on the corresponding test results.

Rounds depend on how satisfactory the performance of the scale measure is.

6. Validate the conceptual framework

Both pre-survey and formal survey are processes to validate the conceptual framework. Evaluation overview

The applicability of a conceptual framework is primarily based on its measurement properties, including reliability and validity.

(1) Reliability: Reliability refers to the reliability of measurement results obtained under similar conditions.

Consistency is used to evaluate the reliability of a measurement tool. Commonly used reliabilities of PRO scales

The indicators include test-retest reliability, internal consistency reliability and intra-tester reliability. test retest letter

The degree is used to evaluate the repeatability of the scale. The correlation coefficient between the initial test and the retest is different.

It should be too low. Internal consistency reliability is used to evaluate the internal consistency of a scale. It is commonly used

Cronbach's ÿ coefficient evaluation (usually no less than 0.7 is appropriate). Tester's internal message

The degree is usually measured using the intraclass correlation coefficient (ICC).

Evaluation, some literature reports believe that ICC consistency can be divided into <0.4 as poor, 0.4-

0.75 is fair, >0.75 is very good.

(2) Validity: Validity refers to the extent to which a measurement reflects the intended measurement

The content is used to evaluate the validity of the measurement tool. A good scale should both

Reliable and effective. High reliability does not necessarily mean high validity (e.g. major depressive disorder

Symptom scale has high reliability and validity for measuring major depression, while it has high reliability and validity for measuring major depressive disorder.

may have high reliability but low validity when measuring mania), but if reliability is low, validity must

Of course it won't be high.

There are many methods to evaluate scale validity, and the 3C method is more commonly used, that is, the internal

content validity, criterion validity, and

Construct validity. Content validity is primarily based on expert knowledge

and experience as well as the patient's subjective judgment whether the dimensions and items of the scale are consistent with

whether it can correctly reflect the content you want to measure. Criterion validity represents the amount of research and development

How well the scale correlates with the so-called "gold standard" scale. Since a gold standard usually does not exist,

And if it exists, the R&D significance is limited (only the R&D scale is of great convenience

sexual situations), so it is less commonly used. Construct validity is often measured through exploratory and experimental

Structural and conceptual framework generated by confirmatory factor analysis method to evaluate observational data

consistency.

In addition to the above 3C concepts, another important indicator of validity is the detection of changes.

The ability to change, also known as responsiveness, is the ability to sensitively reflect changes in patient outcomes (such as

Changes before and after intervention, responses to different interventions, etc.).

7. Write the scale instructions

In order to ensure the correct use of the scale, instructions for use of the scale should be written. quantity

The description of the form includes but is not limited to: target group, complete information including introduction words

Table structure, assignment of dimensions and items and scoring rules for scales, measurement performance,

Provisions for valid responses, handling of missing data, recall periods (if involved), etc.

(2) Translation and/or cultural adaptation of patient-reported outcome measurement scales

If the original scale used for PRO measurement in clinical research is in a foreign language, usually

It often needs to be translated into Chinese before it can be used. One or several items of the original research scale

If the patient cannot understand or obtain the information due to cultural differences,

There are also issues of cultural adjustment involved when working together effectively. Translation and/or culture of the scale

Whether the adaptation is appropriate depends on whether the translated and/or culturally adjusted scale is the same as the original research scale.

Whether the measurement performance is similar is the criterion. Translation and/or cultural adaptation of scales

You can proceed as follows:

1. Preparation stage. Review all relevant information on scale development; organize multi-disciplinary

Translation team (such as English to Chinese, Chinese to English, medical and other professionals); establish and

Communication channels for scale developers, in addition to obtaining authorization to use the latest version of the scale

Except permission, we can better understand the meaning of the scale through communication so that the translation can be more accurate.

Indeed.

2. Forward translation. Two or more translators independently translate the original language version of the

The scale was translated into a Chinese version, and then the various translation manuscripts were combined to form the first Chinese draft.

3. Back translation. Translators who are native speakers of the original language and are familiar with Chinese

Translate the first draft of the article back to the original language, and compare the back-translated version with the original text. If there is any discrepancy,

There is a big difference, and the Chinese translation needs to be further modified until the difference between the back-translation and the original text is

Reach an acceptable level and form the first version of the Chinese version.

4. Cultural adaptation of the scale. If individual items in the scale are inappropriate for the local area

Culture needs to be adjusted. Whether the adjustment results are satisfactory should be measured on the post-adjustment scale.

The judgment principle is that the measurement performance is similar to the original research version.

5. Chinese version first version test. Using the Chinese version to conduct patient assessments in the target population

Conduct a cognitive interview to evaluate the understandability of the scale items and the patient's cognitive level

etc., and conduct quantitative testing of the performance of the scale. If the measurement performance of the scale is consistent with the original

If the research version is similar, the Chinese version can be finalized; if the difference is large, further refinement is required.

Improve the Chinese version until the measurement performance meets the requirements and form the final Chinese version.

6. Chinese version of R&D report. After the final version of the Chinese version is formed, the R&D report is completed.

Report, record the entire R&D process, report measurement performance, and write scale instructions.

Declare the Chinese version of the software copyright when necessary.

(3) Improvement of patient-reported outcome measurement scales

When existing scales are not completely suitable for the research project, they should be improved and used.

For example, after analyzing data from early clinical trials (such as phase II), the scale used is not satisfactory

To meet the required reliability and/or validity for research, the scale needs to be improved or a new one needs to be developed.

scale. Before conducting phase III trials, the scale should be tested again to ensure that

Ensure that the scale used in Phase III trials has sufficient reliability and validity.

4. Selection and evaluation of patient-reported outcome measurement scales

As a PRO measurement tool, the scale should have good measurement performance and should be able to

Reliable and effective. Choose the right PRO test tool for your proposed research project

The quantitative scale is very critical. Combining scientificity and operability, it is recommended to focus on

Key points below:

1. Applicability of the scale: Examine the construction of the scale and pay attention to whether its overall concept

meet the purpose of scale development and fit the applicable population. The research target population should be consistent with the original

The applicable population of the research scale is consistent.

2. Standardized documents or systems: Are there standardized scale-related documents or systems?

Including but not limited to documentation (especially explanations of scale scores), user instructions

User manual, standard format for data collection, important reference data (for design

sample size estimation) etc.

3. Research and development process: Is the purpose of using the scale clearly defined? What is the development process?

Is it strictly standardized? Is the structure of the scale (dimensions, items and their weighting) reasonable?

Are the published results exhaustive?

4. Authoritativeness: Whether the research and development results are publicly published in peer-reviewed journals, whether

It has been widely cited and applied, and whether it is recommended by the guideline.

5. Language and culture: Validity verification of the scale takes into account different teaching methods

education, culture and ethnic background; whether the new language version has been translated and

Back translation and verification. The measurement performance of translated and/or culturally adapted scales should be

It is similar to the original research scale.

6. Validation: Whether it is verified through a large enough sample size, item design and

Whether the assignment is reasonable and whether it has sufficient reliability and validity.

7. Feasibility: The feasibility of the scale when used, including but not limited to implementation

operability of the process, overlapping items when using multiple scales, etc. patient's

Excessive response burden can lead to an increase in missing and rejected responses, reducing PRO

Data quality. Factors that increase patient response burden include: too much content in the scale,

The content is highly repetitive, multiple scales are selected at the same time, and some/some of the scales are not meaningful.

The scale interface is large and difficult to read. The items involve privacy issues that are difficult to answer.

Unreasonable project design, etc.

5. Considerations for using patient-reported outcomes in clinical research

(1) Estimation target framework

Criteria and methods for estimating target framework construction proposed in ICH E9 (R1)

The same applies to clinical studies with PRO as the trial endpoint. Estimate target box

The framework needs to be clearly defined in the protocol and statistical analysis plan.

(2) Select patient-reported outcomes as clinical research endpoints

Clinical studies that select patient-reported outcomes as primary or key secondary outcomes

point, the reasons and basis for the selection should be explained, combined with the research purpose and target indications.

The disease mechanism, drug action mechanism, clinical positioning and other factors should be comprehensively considered. right

When using PRO as the primary or key secondary endpoint, the following issues should be noted: ÿ Required

There must be sufficient basis and be consistent with the research purpose; ÿ If the research design fails

Blinding patients will create a greater risk of subjective evaluation bias and should be extremely cautious;

ÿThe observation period should be long enough to reflect clinically significant changes in PRO; ÿIt should

Control the overall Type I error rate; ÿThe sample size determination should fully consider the expected differences

It should be at least clinically significant.

Patient-reported outcomes should be selected to reflect the patient's perception of the drug's effects

by. Drug effects are not limited to effectiveness, but are also reflected in safety, tolerability or

In terms of impact on quality of life and other aspects, rational selection of patient-reported outcomes can help

Research better reflects patient experience, enabling drug development to follow a patient-centered approach

idea.

(3) Explanation of relevant scales in the research plan and research report

Patient-reported outcomes measured using scales as primary endpoints or key secondary

When endpoints are required, they should be described in the study protocol, including but not limited to:

The rationality of selecting and using the scale; if necessary, briefly introduce the development and application of the scale

usage, especially for some scales that are rarely used; how the scale measures performance

Evaluation methods and indicators; collection and quality control of scale data;

Analysis methods; detailed instructions and training plans for the use of scales, etc.

The clinical research report should include but not be limited to: collection of scale data

(valid responses, missing, etc.); report the metric performance of the scale used (e.g., letter

validity), and compare it with the original research scale. When the difference is relatively large, it should be analyzed

Specific reasons and potential impact of comments on study conclusions; detailed analysis of scale data

analysis results and corresponding reasonable explanations.

(4) Effective responses to the scale

Patients may have missing or negative responses when filling out the questionnaire (such as in 5

(a certain level is fixedly checked in responses to level Likert items), which makes

The data of the scale is distorted. Therefore, the use of the scale should set the standard for effective response.

accurate and specified in the scale's instruction manual. For example, if a scale stipulates that more than 15%

(Different scales have different definitions) items are not answered, or all items are checked

A certain level (e.g., "very satisfied") is considered an invalid response for the subject.

Judgment of valid effects needs to be spelled out in the study protocol and/or statistical analysis plan

Answer criteria and explain reasons. If it is ultimately judged to be an invalid response, it will be the same as no response

Treat as missing values. In some cases, in addition to considering whether the entire scale

In addition to valid responses, a certain dimension of the scale may be regarded as a key variable. In this case,

Whether the dimension response is valid can be specified in advance.

(5) Missing data

It is common for PRO data, especially data measured by scales, to be missing.

Therefore, quality control should be strengthened during the implementation of the study to reduce defects as much as possible.

Very necessary. For missing item data in multidimensional scales, filling in

Complementary method, the specific method is preferred to the method provided in the original research scale instructions, and the

This time, mainstream methods in literature reports were used, and again through the exploration of current research data.

Determined by exploratory analysis (usually done in exploratory research). If no padding is done,

In addition to having too many missing responses that are considered invalid responses, it is necessary to follow the provisions of the original research scale.

Or define in advance in the scheme how to process the entire scale and

Rules for the scores of each dimension. should be developed during the trial design phase to account for missing data

Sound statistical analysis strategies.

(6) Multiplicity issues

When a PRO is listed as one of the primary endpoints or a key secondary endpoint, it will involve

and multiplicity issues. For general handling principles, please refer to "Multiplicity in Drug Clinical Trials".

Issue Guiding Principles (Trial)". Sponsors need to prepare clinical research protocols and systems

Decision-making strategies for multiple problems are specified in the analysis plan in advance

and multiplicity adjustment methods. Scales used by PROs often include multiple dimensions, e.g.

If one or several of these dimensions have important clinical significance and are included in the plan

Listed as a key secondary outcome (the sponsor intends to claim this specific benefit in the package insert),

Multiplicity issues will also be involved, and the control of the overall Type I error rate needs to be considered during design.

Due to the multi-dimensional and multi-item nature of the scale, in addition to focusing on the overall scale

In addition to the analysis of scores, analysis of various dimensions and items is also necessary.

are concerned with multiplexing issues, as long as they are not listed as primary endpoints or

Key secondary endpoints, or no specific benefit claimed in the labeling, do not need to be studied

Perform multiplicity adjustments.

(7) Interpretation of results

Interpretation of results from scale-based PROs and other methods used to assess treatment benefit

The endpoint indicators are the same, and the positive results must have both clinical significance and statistical significance.

minimum clinically important difference,

MCID) is often used to define a threshold of clinical significance, e.g., using a 10-point scale

When measuring the pain level using the Sensory Analog Pain Scale, the average score dropped significantly before and after the intervention.

Less is clinically significant, or the difference between the two groups in the mean decline from baseline is greater than more

Few are clinically significant. When determining MCID, relevant guidelines and experts should be preferred

Consensus and other recognized standards; if there is no recognized standard, you need to contact the regulatory agency

Communicate and reach consensus in a timely manner, statistical methods may provide certain basis for it

according to.

Use statistical methods to estimate MCID. Commonly used methods include distribution-based methods and

Anchoring based approach. The anchoring method is more reliable and facilitates cross-examination across different experiments.

comparison, which sets an external big picture based on the patient's perception of clinical significance

indicators (such as no improvement, slight improvement, significant improvement), and then determine the corresponding

The amount of change in scale scores. Typically, global indicators (level variables) are related to quantities

The correlation coefficient of the change in table score must be at least 0.3 to be meaningful. Some studies

Research suggests that a correlation coefficient of 0.3 is low correlation, and 0.5 is high correlation. Estimate MCID

There are other statistical methods, such as methods based on mixed linear models, which can be combined with monitoring

The main methods will be determined after communicating with the management agency.

(8) Quality control of PRO/ePRO

The number of different research centers, patients, and observers during the study implementation process should be ensured.

The consistency of data collection improves the quality of clinical research. In the plan, at least

To be clear but not limited to:

- Establish quality control standard operating procedures;
- Timing and sequence of PRO/ePRO data collection;
- Training and guidance for relevant personnel on the use of PRO/ePRO measurement tools,

Including methods and standards for judging the completeness of the scale, data filling, storage and transmission

time and method, etc., so that they can fully understand the purpose of using the scale and the description of the scale.

The specific content of the book and the quality control aspects of the scale data collection process;

- Data Management Plan for PRO/ePRO.

In addition, clinical studies using PRO/ePRO require more continuous and proactive implementation

On-site monitoring to ensure the integrity and accuracy of PRO/ePRO data collection.

(9) Use of PRO/ePRO in real-world research

In real-world research, PRO/ePRO is mostly used in prospective studies.

Such as prospective observational studies or practical clinical trials. Number of PRO/ePRO collected

For specific methods of data management or governance, see "Truth for Producing Real-World Evidence"
Real World Data Guiding Principles (Trial)".

6. Electronic patient-reported outcomes

(1) ePRO measurement

Compared with paper PRO, ePRO is more efficient in data collection, real-time,
It has obvious advantages in flexibility, compliance, security and patient privacy protection.
potential. The shortcomings of ePRO are mainly reflected in the fact that some patients may have difficulty operating electronic equipment.
encounter difficulties in equipment, especially the elderly, young, and those who have limited movement due to illness.
Patient population with manual manipulation ability.

Currently, the collection of ePRO data generally includes telephone-based interactive voice response.
There are two types: answering system and screen-based reporting system. phone-based interactive
Voice response systems feature automated calls using pre-recorded questions and
Answer options script and allow patients to record responses using keystrokes, with data stored directly
stored in the central database. Screen-based reporting system can be installed on the patient's own
on electronic devices such as smartphones, tablets, computers, and even wearable
Wearing a medical device, also known as BYOD, allows patients to access a website on the device or
The software selects answers based on its own situation and records are saved.

The ePRO system can be connected to electronic medical record systems or electronic data collection systems.
Form a complete data flow at the individual level; its time recording function can effectively prevent and
Identify behaviors that affect data reliability such as response backfilling or early responses; their remote
The monitoring function helps researchers and data managers conduct real-time online data management.

processing and remote data monitoring, questioning and annotating questionable data, and promptly correcting

Subjects were interviewed.

(2) General considerations for using ePRO

ePRO measurement tools and data in clinical studies for drug registration purposes

Collection and data management, etc., should follow drug clinical trial data management, electronic data

Basic requirements for data collection and real-world data governance related guiding principles.

The ePRO measurement method based on the network platform is different from the paper PRO measurement

Tools whose data are usually uploaded to online data collection centers for comprehensive management by users

management to realize data storage, monitoring and export. Therefore, to ensure that researchers have

Electronic source data maintenance and preservation rights, the research institution has original document support to

For sponsor audit and regulatory department verification, the following should be followed when using the ePRO measurement tool

The following principles:

1. Researchers should have the ability to maintain and confirm the accuracy and authenticity of ePRO source data.

sexual authority. Researchers capture ePRO data through audit trails and measure equipment

Prepare any data changes and modifications after uploading to avoid the sponsor or third-party agencies

Have sole control over the collection/management system of raw ePRO data. ePRO source data is

Refers to the record originally recorded by the ePRO system and stored in the database, if

The initial records of the ePRO system are directly imported into the EDC system and stored in eCRF, then

The initial eCRF is the source data.

2. Data security management system and access control mechanism. Use encryption technology to protect

Ensure the integrity, confidentiality, and confidentiality of data during collection, extraction, transmission, and storage.

Traceability prevents any individual or institution from modifying original data or deleting patient reports

Reported adverse events, high-risk warnings and other data, and establish corresponding access control mechanisms

system to avoid the risk of unplanned unblinding.

3. Data backup. Avoid data damage or loss, loss of data during the test

The risk of being unable to reconstruct or verify the source data.

4. Data saving. Research institutions and researchers should maintain electronic source data or

Electronic documents allow regulatory inspectors to inspect,

Verification and reproduction of original data.

If analysis of research data reveals that ePRO scale measurement performance is

There is a big gap between the research scales and the potential problems in the implementation of the ePRO scale should be considered.

problem and correct it. In addition, the ePRO measurement tool based on item response theory,

Using computer adaptive testing technology, select the following item based on the answer to the previous item

items, thereby reducing the number of items to reduce patient response burden, but reducing the number of items

The quantity should meet the premise of ensuring the validity of the content of the scale. Sponsors use this class

ePRO measurement tool, you need to submit conceptual framework construction and item library design and screening process

procedures, program construction rules, and result analysis and interpretation and other related information.

7. Communication with regulatory agencies

When a sponsor plans to use PRO/ePRO as a primary or primary confirmatory study

When determining secondary endpoints, timely communication should be made with regulatory agencies. Communication issues include but

Not limited to target indication disease background, select PRO as primary or key secondary

Reasons and basis for study endpoints, type of study design, and development scale (if any)

Validation conceptual framework and scale instructions and other materials, PRO/ePRO improvements and/or

or cultural adaptation (if any) and basis, verification of reliability and validity, minimal clinical

Differences in meaning and basis, quality control in implementation and other issues. Before communicating,

Sponsors should provide the regulatory agency with a preliminary report containing PRO/ePRO statistical analysis

Considered trial protocols and PRO/ePRO related information. During the test, if

Major adjustments to the clinical trial protocol due to changes to PRO/ePRO should be

Regulatory agencies communicate in a timely manner.

references

[1] Acquadro C, Berzon R, Dubois D, et al. Incorporating the patient's perspective into drug development and communication: an ad hoc task force report of the Patient-Reported Outcomes (PRO) Harmonization Group meeting at the Food and Drug Administration, February 16, 2001.Value Health. 2003; 6(5): 522-531.

[2] Bukhari M. “PROMs vs. PREMs (Patient-Reported Experience Measures).”; Patient Reported Outcome Measures in Rheumatic Diseases. Ed. Miedany YE. London: Springer, 2016; 405-417

[3] Byrom B, Watson C, Doll H, et al. Selection of and Evidentiary Considerations for Wearable Devices and Their Measurements for Use in Regulatory Decision Making: Recommendations from the ePRO Consortium. Value Health. 2018; 21(6): 631-639.

[4] Calvert M, Blazeby J, Altman DG, et al, CONSORT PRO Group. Reporting of patient-reported outcomes in randomized trials: the CONSORT PRO extension. JAMA 2013; 27; 309(8): 814-822.

[5] Cohen J. A power primer. Psychological Bulletin 1992;

112(1): 155–159

[6] Coons SJ. ePRO systems validation: clearly defining the roles of clinical trial teams and ePRO system providers. Value Health. 2013; 16(4): 457-458.

[7] Coons SJ, Gwaltney CJ, Hays RD, et al. Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (PRO) measures: ISPOR ePRO Good Research Practices Task Force report. Value Health. 2009; 12(4): 419-29.

[8] Copay AG, Subach BR, Glassman SD, et al. Understanding the minimum clinically important difference: a review of concepts and methods. Spine J Off J North Am Spine Soc. 2007; 7: 541–546.

[9] Doward LC, Gnanasakthy A, Baker MG. Patient reported outcomes: looking beyond the label claim. Health Qual Life Outcomes. 2010; 8: 89.

[10] EMA. Reflection paper on the regulatory guidance for the use of health relate quality of life (HRQL) measures in the evaluation of medicinal products. 2005.

[11] EMA. Reflection paper on expectations for electronic source

data. 2010.

[12] EMA. Reflection paper on the use of patient reported outcome (PRO) measures in oncology studies. 2014.

[13] Fayers PM, Machin D. Quality of Life: The assessment, analysis and reporting of patient-reported outcomes (3rd Edit). John Wiley & Sons, Ltd. 2016.

[14] Ferreira ML, Herbert RD, Ferreira PH, et al. A critical review of methods used to determine the smallest worthwhile effect of interventions for low back pain. J Clin Epidemiol. 2012; 65: 253–261.

[15] FDA. Clinical outcome assessment (COA) compendium. 2021.

[16] FDA. Clinical outcome assessment (COA) qualification program. https://www.fda.gov/drugs/drug-development-tool-ddt-qualification-programs/clinical-outcome-assessment-coa-qualification-program.

[17] FDA. Guidance for industry: Assessing COVID-19-Related Symptoms in Outpatient Adult and Adolescent Subjects in Clinical Trials of Drugs and Biological Products for COVID-19 Prevention or Treatment. 2020

[18] FDA. Guidance for industry: Electronic source data in

clinical investigations. 2013.

[19] FDA. Guidance for industry: Patient-Reported Outcome Measures: use in medical product development to support labeling claims. 2009.

[20] FDA. Patient-Focused Drug Development: collecting comprehensive and representative input. 2020.

[21] FDA. Roadmap to patient-focused outcome measurement in clinical trials. 2015. https://www.fda.gov/media/87004 /download

[22] FDA. Plan for issuance of patient-focused drug development guidance. 2017.

[23] FDA. Upper facial lines: developing botulinum toxin drug products. 2014.

[24] Fiero MH, Pe M, Weinstock C, et al. Demystifying the estimand framework: a case study using patient-reported outcomes in oncology. Lancet Oncol 2020; 21: e488–94

[25] Fleiss JL. Measuring agreement between two judges on the presence or absence of a trait. Biometrics, 1975ÿ31: 651-659.

[26] Fox MW, Onofrio BM, Onofrio BM, et al. Clinical outcomes and radiological instability following decompressive lumbar laminectomy for degenerative spinal stenosis: a comparison

of patients undergoing concomitant arthrodesis versus decompression alone. J Neurosurg. 1996; 85(5):793-802.

[27] Hong K, Majercak KR, Villalonga-Olives E, et al. Patient-reported outcomes in breast cancer FDA drug labels and review documents. J Patient Rep Outcomes. 2021; 5(1):36.

[28] Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Control Clin Trials. 1989;10:407–415.

[29]Lawrance R, Degtyarev E, Griffiths P, et al. What is an estimand& how does it relate to quantifying the effect of treatment on patient-reported quality of life outcomes in clinical trials? J Patient Rep Outcomes. 2020; 4(1):68.

[30] Ly JJ, Crescioni M, Eremenco S, et al. Training on the use of technology to collect patient reported outcome data electronically in clinical trials: best practice recommendations from the ePRO Consortium. Ther Innov Regul Sci. 2019; 53(4): 431-440.

[31] Mokkink LB, Terwee CB, Knoll DL, et al. Protocol of the COSMIN study: COnsensus-based Standards for the selection of health Measurement Instruments. BMC Med Res Methodol. 2006; 6: 2.

[32] Walters S. Quality of life outcomes in clinical trials and health-care evaluation: A practical guide to analysis and interpretation. John Wiley & Sons, Ltd. 2009.

[33] Wild D, Grove A, Martin M, et al. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: Report of the ISPOR task force for translation and cultural adaptation. Value Health. 2005; 8(2): 94-104.

[34]Guideline, ICH. "Addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials." E9 (R1). Step 4 (2019): 20.

Appendix 1: Glossary

Criterion Validity: also known as calibration validity , refers to the developed PRO

Scores on the scale are consistent with known so-called “gold standard” scales that measure the same concept.

degree of concern. Most PRO scales cannot measure their criterion validity because there is no gold standard.

Instrument: A tool that captures data and supports its use

Tools (e.g. scales) with information and documentation, often including detailed instructions for implementation

citations, standard formats for data collection, description of scoring and analysis methods, and objectives

Normative documents for interpretation of results for target disease populations, etc.

Concept : also known as concept of interest (COI).

At the supervisory level, the concept is that the PRO scale captures or reflects the individual's clinical,

Biological, physiological, functional, etc. state or experience. At the PRO level, the concept

Represents a patient's functioning or feelings about their health condition or related to treatment.

Patient **-reported Outcome** , PRO): for anyone who comes

Information about one’s own disease and related issues reported directly by the patient and not modified or interpreted by others

Outcomes should be assessed for perceived treatment.

Patient-focused Drug Development

PFDD): refers to a systematic approach that includes

Help ensure patients’ experiences, perspectives, needs and priorities are captured and

Effectively integrate into drug development and evaluation.

Recall Period **(Recall** Period): The patient’s response time is

PRO entry or question time. Memories can be instantaneous (real time), or

A retrospective from some time ago. The recall period should not be too long, usually no more than a week.

Ability to Detect Change : Measurement tool detection

PRO measurement scores vary with measurement conditions (different time points before and after intervention, different interventions,

The ability to show differences due to changes in different groups of people, etc.).

Construct Validity : Also known as construct validity, it refers to the

According to the presented PRO scale items, dimensions and concepts to be expressed

Whether the structural relationships are consistent with the theoretical constructs for scale development.

Cronbach's alpha coefficient: a reliability index used to evaluate the internal consistency of the scale.

Conceptual Framework **of a Scale: Basic**

Constructed from previous research (documentation), expert knowledge and experience, and necessary research

The dimensions and items of the scale were constructed. Study on the number and naming basis of dimensions

The understanding of the research content is set, and the number of items and the content of the items under each dimension are used to understand

reflect the connotation and importance of the dimension to which it belongs (for example, when each item is equally weighted,

The number of entries under a dimension reflects the importance of the dimension).

Content Validity : Qualitative research based on expert knowledge, verification

Verify that the scale can measure what it is intended to measure.

Quality **of Life (QoL):** also known as quality of life, quality of life,

Used to assess overall health as reflected in all aspects of life.

Item : A question, statement, or task (and standardized response options

items), are used in patients' assessments of specific concepts.

Adaptation : Consideration based on language and cultural differences between ethnic groups

Any changes made to the table. Adaptation does not change the structure of the PRO scale, but it

Adapt a small portion of the content to another mode, language, or group of people. Adaptation research

It is to verify the measurement performance of PRO scale in new environment or new language.

Dimensions **(Domains/Dimensions/Factors): The first level structure that constitutes the scale**

structure (secondary structure scale) or first and second level structures (tertiary structure scale),

Used to express a certain aspect (concept) that makes up the scale. A dimension consists of a

or multiple entries.

Validity : refers to the extent to which a measurement reflects the content it is intended to measure.

content, used to evaluate the validity of the PRO scale.

Reliability : refers to the consistency of measurement results obtained under similar conditions.

Consistency and repeatability are used to evaluate the reliability of the PRO scale.

Symptom : a disease or health condition that can only be noticed and perceived by the patient

or any subjective evidence of therapeutic efficacy.

Treatment Benefit: The impact of treatment on a patient's survival, feeling, or function

energy impact. Treatment benefit can be demonstrated by efficacy or safety advantages. example

For example, treatment effectiveness can be measured by improvement or delay in symptom progression, or

Can be measured by reducing or delaying treatment-related toxicity. No direct access to treatment

Measures of the effect on a patient's survival, feeling, or function are surrogate measures of treatment benefit

quantity.

Minimum Clinical **Important Difference,**

MCID): a threshold often used to define clinical significance, e.g., using a 10-point scale

When measuring pain level using the Visual Analog Pain Scale (VAS), the average before and after intervention

What is the level of decrease in scores that is clinically significant, or the average decrease in scores from baseline between the two groups?

The difference must be greater than what is clinically significant.

Appendix 2: Comparison of Chinese and English vocabulary

Chinese	English
criterion validity	Criterion Validity
Measuring tools	Instrument
Intra-tester reliability	Intra-rater Reliability
test-retest reliability	Test-retest Reliability
Electronic Patient Reported Outcomes	Electronic Patient-reported Outcome (ePRO)
electronic data capture	Electronic Data Capture (EDC)
Measuring performance	Measurement Properties
conceptual framework	Conceptual Framework
patient reported outcomes	Patient-reported Outcome (PRO)
Patient-centered drug development	Patient-focused Drug Development (PFDD)
recall period	Recall Period
scale	Scale
Summated Xerostomia Inventory , (SXI)	
interactive voice response system	Interactive Voice Response Systems (IVRS)
Screen -based Reporting Devices	
Anchoring based approach	Anchor-based Method
Ability to detect changes	Ability to Detect Change

Chinese	English
construct validity	Construct Validity
Internal Consistency ReliabilityInternal Consistency Reliability	
content validity	Content Validity
Quality of Life	Quality of Life
visual analog scale	Visual Analog Scale (VAS)
entry	Items
entry pool	Item Pool
debug	Adaptation
Dimensions	Domains/Dimensions/Factors
item response theory	Item Response Theory (IRT)
validity	Validity
reliability	Reliability
response burden	Respondent Burden
intraclass correlation coefficient	Intraclass Correlation Coefficient (ICC)
symptoms	Symptom
major depressive disorder symptom scale	Symptoms of Major Depressive Disorder Scale (SMDDS)
treatment benefit	Treatment Benefit
Bring your own equipment	Bring-Your-Own-Device (BYOD)
minimal clinically meaningful difference	Minimum Clinical Important Difference (MCID)