Summary HTA

DIMDI

german agency for HTA of
German Institute of Medical
Documentation and Information

HTA-Report | Summary

Informative value of Patient Reported Outcomes (PRO) in Health Technology Assessment (HTA)

Brettschneider C, Lühmann D, Raspe H

Health political background

In Germany the legal framework of the SGB V puts the evaluation of the "medical benefit" in the centre of a technology assessment, although the term "benefit" is not defined by legislation. However, the wording of § 27 suggests, that the requirements of SGB V can be fulfilled by reporting a technology's effects on morbidity, mortality and quality of life. Further characterisations of these endpoints are found in the "Methodenpapier 3.0 (05/27/2008)" of the Institute for Quality and Efficiency in Health Care (IQWiG). In chapter 3.1. the expression "patient-relevant" is defined as "… how a patient feels, is able to function and participate or survive." Aside from morbidity, mortality and quality of life, intervention- or disease-specific efforts and patients' satisfaction with treatment may constitute secondary endpoints. It is mandatory, that all kinds of endpoints directly and reliably indicate changes in health status. Finally, the only way to validly assess patients' quality of life, some aspects of morbidity and patient satisfaction is by direct questioning of the patients – that is why they are called Patient-Reported Outcomes (PRO).

Scientific background

PRO is used as an umbrella term for different concepts aiming at the measurement of subjectively perceived health status e. g. as treatment effects. Their common characteristic is, that the appraisal of the health status is reported by the patient himself. The American Food and Drug Administration (FDA) defines PRO as follows: "A PRO is any report of the status of a patient's health condition that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else". PRO are used if a concept may be assessed best by the patient himself. PRO may be elicited either as a single value (e. g. severity of pain) or as a value of change between two measurements (e. g. pain cessation).

The measurement of PRO is based on two different approaches. The psychometrical approach refers to the reporting of perceived symptoms (e. g. their presence and severity), capabilities, behaviours, or emotional or mental state. Single dimensions may be summarised forming complex concepts (e. g. quality of life). The preference-based approach refers to the value a patient assigns to a specific health status. Methodological approaches to elicit preferences are based on econometrics and decision theory. The following text exclusively refers to the psychometrical approach.

Aside from health economics, there are two main applications for PRO measurements: as indicators for quality management in health care and as an endpoint in clinical trials. Assessing the benefit of a health care technology on the basis of published clinical trials is the main goal of Health

Bundesministerium
für Gesundheit

Technology Assessment (HTA). Therefore, determining the role of PRO in HTA will focus on their application in clinical trials.

Psychometric properties such as sufficient validity (content, construct, and criterion), reliability and responsiveness are basic requirements for instruments measuring PRO in clinical trials. Other aspects are administrative and economic feasibility and acceptability.

In order to outline the informative value of PRO in HTA, first of all an overview of concepts, classifi¬cations and methods of measurement for PRO is needed. Strengths and weaknesses of the different concepts and measurement approaches need to be pointed out, especially with regard to their psychometric properties and interpretability of results. The overview is based on a systematic analysis of publications that deal with the theoretical framework for the evaluation of PRO.

The current role of PRO measurements in HTA is outlined by two empirical analyses. The first one determines the frequency, the type and consequences of PRO measurement in randomized controlled trials (RCT). Treatment of rheumatoid arthritis with biologicals and chemotherapy of breast cancer were chosen as examples. These indications – treatment pairs were selected in order to represent one chronic disease and one life-threatening acute disease.

The second empirical analysis focuses on the role of PRO in the work of HTA-agencies. HTA-reports on the two example conditions are analysed to determine to which extent conclusions of the reports are based on PRO. Furthermore, method guides of international HTA-agencies are screened in order to determine the role that these agencies attribute to PRO measurements. These overarching objectives are broken down into eight research questions.

Research questions

Theoretical background

1. How are PRO defined in the context of clinical trials?
2. What are the classification systems to characterise endpoints for clinical trials and how are PRO fitted into these systems?
3. What are the methodological approaches for PRO measurements, what are their specific strengths and weaknesses?

Empirical analyses

1. How frequently are PRO chosen and reported as primary and/or secondary endpoints in RCT analysing the effects of antibody treatment of rheumatoid arthritis and chemotherapy of breast cancer?
2. Which PRO are utilized in these trials?
3. Do conclusions on the effectiveness of treatments based on PRO match with the conclusions based on clinical or radiological endpoints?
4. As concerns assessing the health benefit of medical interventions, what is the informative value attributed to PRO by the methodological literature, especially methods guides of HTA-agencies?
5. Do HTA-reports on rheumatoid arthritis and breast cancer refer to PRO results – especially in their conclusions and recommendations?

Methods

Answering the research questions requires three different literature search strategies.

The first one aims at identifying publications, which report on definition, classification and operationalisation of PRO. It covers the publication period from 1990 to 2009 and is performed in the medical databases of the German Institute of Medical Documentation and Information (DIMDI). The selection of relevant publications is executed by a two-step approach based on defined inclusion and exclusion criteria. A standardised extraction of data employing extraction sheets is impossible, because of the narrative nature of the publications. The extraction of relevant information is performed by matching the publications to a research question in a first step and then by matching relevant information to categories derived from the literature.

The second search strategy aims at identifying RCT that examine the efficacy of biologicals for the treatment of rheumatoid arthritis and chemotherapy for the treatment of breast cancer. Again the search is performed in the medical databases of DIMDI. It covers the publication period from 2005 to 2009. The two-step selection procedure is based on two sets of inclusion and exclusion criteria. Methodological quality of RCT is assessed by a modified version of the Cochrane Risk of Bias Tool. All relevant data are extracted into extraction sheets.

The third search strategy is constructed to retrieve methodological guidance documents from HTA-agencies and institutions of pharmaceutical evaluation as well as HTA-reports on rheumatoid arthritis and breast cancer assessing the treatments specified above. Methodological guidance documents are searched for by hand on the websites of the INAHTA member agencies and institutions (INAHTA = International Network of Agencies for Health Technology Assessment), which are identified by the HTA-report "Methods for the comparative evaluation of pharmaceuticals". HTA-reports on the two example topics are searched for in the HTA-database of the Centre for Review and Dissemination. The selection of the literature is based on inclusion and exclusion criteria, data extraction is performed into extraction sheets.

Results

After the selection 158 relevant publications remained to answer the methodological research questions (1 to 3). Furthermore, 129 RCT reporting results of chemotherapy in breast cancer patients and 59 RCT reporting results of treatment of rheumatoid arthritis with biologicals are included to answer the empirical research questions related to clinical trials (4 to 6). To answer the empirical questions concerning HTA-reports (7, 8) 34 methodological guidance documents, twelve reports on chemotherapy of breast cancer and nine reports on biological therapy of rheumatoid arthritis are included.

Theoretical issues

In its current meaning the term PRO is used since about the year 2000. Its definition is pragmatic and refers to the source of information. Essential aspects of a PRO are the patient report and his or her prerogative of interpretation of the health status. PRO substituted the term "quality of life" as the umbrella term for endpoints reported by patients themselves. Constructs subsumed under the umbrella term PRO are perceived symptoms, functioning, health perception,

satisfaction and (health related) quality of life. In the methodological literature there is no consensus regarding the definition of these constructs.

Several research groups try to integrate the different constructs into comprehensive models, which display the relationships and interactions between parameters and illustrate external influences. Wilson and Cleary created a linear, five-step model. Its poles represent biological and physiological variables on the one side and overall quality of life on the other side. Symptom status, functional status and general health perceptions bridge the gap between the two poles. Between two neighbouring constructs interdependencies are assumed.

Another comprehensive model is the International Classification of Functioning, Disability and Health (ICF) of the World Health Organisation (WHO). The key intention of the ICF is the description of the health state of an individual with regard to the dimensions body structures, body functions and activities and participation. The health state assigned by the ICF is influenced directly by the underlying disease (classified by International Classification of Diseases (ICD)) and environmental as well as personal factors.

Valderas and Alonso integrate both approaches. Their model is designed to classify instruments measuring patient reported endpoints. Definition and classification take three aspects into consideration: construct, target population and measurement approach.

The major part of the methodological literature provides information related to the operationalisation of PRO. Any PRO measurement should be based on a conceptual framework, which specifies the relevant constructs, the definition of the target population and an endpoint model specifying relevant variables and instruments.

There is a variety of instruments available to measure PRO. Most of them are based on classical test theory, some of them on a newer approach called item response theory. Instruments based on item response theory are shorter and produce more focussed results. Psychometric instruments document the description of a health status from the perspective of the patient. Preference based instruments require the valuation of a health status by the patient. Generic instruments may be applied independently of population characteristics while specific instruments may only be utilized in populations with certain characteristics (e. g. disease specific instruments). Usually results from measurements with specific instruments are more precise than those from generic instruments, while the latter allow comparisons across patient populations with different diseases. Furthermore, there are profile and index instruments. While profile instruments present separate values for every construct measured the results of index instruments are summarised into one figure.

During the development of new instruments item generation is a pivotal step. According to the literature a mixed approach consisting of literature studies, expert input and focus groups is considered state of the art.

Measurement properties are crucial for the applicability of an instrument. Main properties are validity, reliability and responsiveness. Validity may be determined as content, construct or criterion validity – depending on the type of instrument and the available comparisons. Reliability primarily refers to test-retest-reliability and internal consistency. Responsiveness describes the ability of an instrument to record changes of an endpoint. It is determined by a distribution based or an anchor based approach. The former utilizes effect size while the latter requires comparisons with an external anchor (gold standard).

A pivotal concept for the interpretation of results from PRO measurements is the minimal important difference (MID) – describing the smallest difference between two measurement results that a patient considers relevant. There is no standard approach to determine the MID. The most frequently used approaches correspond with the approaches used to determine responsiveness.

For methodological and economical reasons the use of existing standardised and validated instruments is preferable to the development of new instruments. In some situations it may be necessary to modify an instrument to match a specific context. A particularly critical point is the translation of instruments into a foreign language. Guidelines suggest a ten-step process of forward and backward translations. In any case a modified instrument needs to be revalidated.

PRO in clinical trials

The literature searches retrieved 73 publications reporting on 59 RCT that investigate the effects of treating rheumatoid arthritis with biological drugs. Most frequently the American College of Rheumatology (ACR) core set of disease activity measures is used as the primary endpoint. This composite endpoint consists of seven criteria, among them three PRO (pain, global assessments of health status and functioning). For a positive score, improvements in at least one PRO criterion are required. The 20-percent criterion, which requires an overall 20-percent improvement is used as primary endpoint in 23 studies. Pure PRO are less frequently used as primary endpoints. The most frequently applied PRO is the health assessment questionnaire (HAQ), a questionnaire investigating functional status by 20 items. It is defined as the primary endpoint in six trials. The disease activity score (DAS) is the most frequently studied secondary endpoint (39 trials). The DAS again is a composite endpoint consisting of four criteria, among them one PRO (health perception). Pure PRO are defined secondary endpoints in 34 trials. The dominant constructs are functional status and health perception. The HAQ (32 trials) and the SF-36 (16 trials) are the most frequently applied questionnaires. The SF-36 measures health perception by asking 36 questions across eight domains. Stratified analysis by methodological quality of trials indicates that trial quality does neither predict the use or non-use of PRO nor the type of constructs investigated.

Five trials report contradictory results of clinical and patient reported outcomes. In one trial worsening of radiological endpoints (synovitis, oedema, and erosions) is not reflected in HAQ results. In the other four trials the occurrence of adverse events (fever, headache, asthenia, and infections) is not reflected by the results of the HAQ or SF-36 respectively. These findings are not commented by the authors. In general, PRO results are commented in the discussion parts of most trials that applied PRO measures. Of 35 trials that discuss the comparison of PRO results and results from traditional endpoints 19 authors come to the conclusion that PRO results are supportive. Interdependencies between PRO and traditional endpoints are discussed in twelve studies. Findings are not always coherent, e. g. between progression of radiologically attested joint erosions and HAQ results. Still, frequently interdependency between disease activity and the HAQ results is reported.

15 studies take reference to PRO in their conclusions – mostly to functional status and health perception. This is particularly true for trials where PRO results and results from traditional endpoints are congruent and therefore suggest identical conclusions. In two trials results from PRO (adherence to therapy, quality of life) are used to refine recommendations. Six trials identify the need for further research to optimise the use of PRO. They recommend the generation of long-term data and the

further analysis of interdependencies between clinical and patient reported outcomes.

The literature searches retrieved 129 publications reporting on 123 RCT that investigate the effects of chemotherapy in breast cancer patients. In these trials the most frequently used primary endpoint is survival (66 times in 53 studies). Only six studies utilize PRO as primary endpoint, mostly quality of life. Three of these studies apply the cancer specific EORTC questionnaire. In 22 trials PRO are defined as secondary endpoints. Again, the quality of life questionnaire of the EORTC is the most commonly applied instrument (twelve trials). As for the rheumatology trials, stratified analysis by methodological quality again indicates that trial quality does neither predict the use or non-use of PRO nor the type of constructs investigated.

Of 26 studies, which present results of PRO, 15 report statistically significant differences between treatment groups. One trial detects improvements of quality of life that are not reflected in overall survival and progression free survival. 24 of 28 trials reporting PRO results comment them in the discussion section: 18 trials by comparing PRO and traditional endpoint results, three trials by pointing out inconsistencies. Interdependencies are discussed in six studies of which three find a strong influence of adverse events on quality of life. Nine trials identify the need for further research, especially on how to implement PRO information into routine patient management.

PRO in methodological guidance documents of HTA-agencies

The searches retrieved 34 methodological guidance documents from the websites of 20 HTA-agencies and pharmaceutical regulatory bodies. 18 of these papers address the topic of PRO to a relevant extent. Three aspects are discussed in some detail: The role of PRO in decision making – some agencies equally weigh PRO results and clinical results while others regard results from PRO measurements as supplementary information only. Furthermore, technical aspects of PRO measure¬ment (choice of instrument, psychometric properties) and the role of quality of life measurement in economic evaluation are addressed.

PRO in HTA-reports (rheumatoid arthritis)

Handsearches and database searches identified nine HTA-reports addressing the treatment of rheumatoid arthritis with biological drugs. In their background section seven reports justified the use of PRO for this indication by referring to the disease's impact on functional status and quality of life.

Analyses of the methods section of reports revealed that nearly half of the documents prospectively specified PRO as relevant endpoints for the assessment. Predominantly the composite endpoints ACR20, ACR50 and ACR70 are referred to but single measures of symptoms (pain scale), functioning (HAQ) or quality of life are mentioned as well.

In all reports results of PRO measurements are discussed explicitly, with a special focus on the ACR criteria. Two reports compare the results from traditional endpoints with PRO results. In the majority of reports the results from PRO measurements are regarded insufficient yet for decision making. The authors point out the need for more research in order to bridge the gap.

PRO in HTA-reports (breast cancer)

Twelve HTA-reports were found that assess different types of chemotherapy for breast cancer. In their background section only three of these twelve reports refer to aspects of everyday life that may be impacted by chemotherapy. These reports regard symptom relief and quality of life along with prolongation of life as the main goals of chemotherapy.

As specified in the methods sections, quality of life data and other constructs such as pain, fear, depression and functional status are taken into consideration by five reports. One report refers to PRO for input into cost-utility-analyses.

PRO results, mainly quality of life data, are presented in the results section of four reports. In their discussion section, three reports refer to quality of life data. Of these, two reports discuss the fact that improvements in disease-free survival not necessarily translate into improvements of quality of life. On the other hand the inevitable relation of toxicity and reduced quality of life is pointed out.

Discussion

The results of the presented report may be compromised by some methodological limitations. First, since there is no controlled vocabulary covering the issue of PRO and the range of databases and search period was limited, some theoretical and methodological publications may be missed. Still it may be assumed that key publications would have been found in reference lists of the publications available. Second, some of the theoretical publications may not represent original research. The selection was hampered by the fact, that there is no specific study type that can guide a selection of methodological literature. This problem may be subordinate though, since no quantitative analyses of the theoretical literature have been performed.

In summary, the research questions may be answered as follows:

Currently the term PRO is used as an umbrella term for patient reported endpoints. Quality of life, which has formerly been used as the umbrella term, is now one of several constructs subordinated to PRO (e. g. symptoms, functional status, health perception, quality of life). There is actually no consensus, which other constructs are part of PRO. Most authors find a rigorous separation of traditional outcomes and PRO unreasonable. Models offer an integrated perspective which presents a holistic impression of the disease and its consequence. Furthermore, interdependencies between endpoints may be visualised. The most complex and most pragmatic approach is presented by Valderas and Alonso.

There are different approaches for the measurement of PRO. Psychometric and preference-based approaches, profiles and indices, generic and specific instruments all have their own strengths and weaknesses and must be chosen depending on the context of the research project. Validity, reliability and responsiveness represent the psychometric properties of an instrument. For methodical and economic reasons the use of existing standardised and validated instruments is preferable to the development of new instruments. If it is necessary to modify an existing instrument, a revalidation of the modified version is required. Relating to the interpretation of results the MID is the pivotal aspect. Item generation for new instruments is still based on the principles of the classical test theory. But new approaches, e. g. the item response theory, which provides shorter and more precise instruments, are improving.

The frequency of PRO in RCT relating to rheumatoid arthritis indicates that the informative value of PRO is regarded high. Composite endpoints (ACR criteria, DAS) which combine clinical endpoints and PRO are most frequently used. Pure PRO, especially functioning and health perception (HAQ, SF-36) are often found as secondary endpoints. A sophisticated discussion of the interdependencies of PRO and traditional endpoints, referring to the specific condition is seldom found.

The RCT of breast cancer use PRO only to a limited extent (24 % of studies) and mainly as secondary endpoints. Quality of life is the only construct that is referred to in these trials. Again, an integrating discussion of PRO and traditional endpoints is hardly found.

About half of the retrieved methodological guidance papers address the informative value of PRO for the evaluation of health benefits as well as for economic evaluations. Comprehensive information is presented for measurement instruments. Reports relating to the indications reflect the informative value of PRO in primary studies. It needs to be pointed out that in most reports a need for further research concerning PRO is stated – this implies, that HTA-agencies promote the use of PRO in a proactive manner.