Publication Series
INTERDISCIPLINARY PLATFORM ON BENEFIT ASSESSMENT

Volume 20
March 2025

Which endpoints are patient-relevant?

Springer Medizin

PUBLICATION SERIES INTERDISCIPLINARY PLATFORM ON BENEFIT ASSESSMENT

ALL VOLUMES AVAILABLE ONLINE AT:
HTTPS://WWW.AERZTEZEITUNG.DE/KOOPERATIONEN/PLATTFORM-ZUR-NUTZENBEWERTUNG

Volume 20
March 2025

Content

Goals of the plattform

Since the introduction of AMNOG in 2011, Germany has a well-established and widely accepted „adaptive system" for the assessment of the patient-relevant additional benefit (Health Technology Assessment, HTA). The assessment of the additional benefit by the Federal Joint Committee (G-BA) is the result of expert work based on a law (AMNOG) and procedural and methodical regulations.

The active players on the side of the G-BA and the health insurance funds are classified as scientists, hospital physicians and office-based statutory health insurance physicians, the Medical Service of the Health Funds and employees of the insurance fund administration, but also as patient representatives, however, they act on the basis of their own interests. Value dossiers for new pharmaceuticals, likewise qualified and interest-based, are submitted to the G-BA by the pharmaceutical companies, which serve as the basis for the assessment of the additional benefit.

Because the supply of pharmaceuticals to the population is significantly influenced by the assessment of the additional benefit, it makes sense to provide critical and careful support for the assessment process with a focus on identifying possible faults and counteracting imbalances. The Interdisciplinary Platform on Benefit Assessment set itself the task of supporting the benefit assessment within a small group of experts with the following objectives:

- Discussing the procedures for the assessment of the additional benefit, including in relation to approval of pharmaceuticals,
- Working towards international standards of evidence-based medicine and of health economy being adhered to as well as applied and further developed,
- Determining whether and to what extent patient-relevant additional benefits, in particular in the areas of mortality, morbidity and quality of life, are identified and which methodological problems occur during the process,
- dentifying possible undesirable developments, in particular with regard to supplying patients with new active substances,
- Enabling and holding a constructive dialogue with all players involved in the benefit assessment procedure, e. g. on the further development of the legal framework conditions of AMNOG.

Moreover, the European perspective in HTA of innovative pharmaceuticals was reinforced by the European Commission's proposal for a Regulation on HTA in 2018. Monitoring the conflict between the well-established national assessment and the intended European HTA harmonisation is also a central concern of the platform. The Interdisciplinary Platform would like to make a contribution to ensuring that new active substances are transparently and fairly assessed. According to the Advisory Council, an interdisciplinary dialogue about the results of the assessment and the applied benefit assessment methods is essential. Furthermore, in the benefit assessment process it sees a good opportunity to inform the prescribing physicians of the expected additional benefits of new pharmaceuticals for patients earlier than it was previously the case.

The Interdisciplinary Platform is a result of the discussion process between clinicians and experts. The mutual desire to pool specialist knowledge in the form of interdisciplinary seminars is supported by an open consortium of sponsors. These include AbbVie Deutschland GmbH & Co. KG, DAK Gesundheit, MSD Sharp & Dohme GmbH, Novo Nordisk Pharma GmbH, Roche Pharma AG and Association of Research-Based Pharmaceutical Companies (vfa e.V.).

The Advisory Council of the Interdisciplinary Platform on Benefit Assessment

Relevance of clinical endpoints for patients: status quo and outlook

Professor Jörg Ruof

Dear readers

Yes, it is true: the question of the relevance of individual endpoints of clinical trials for patients is as old as the AMNOG itself - and yet this issue has still not been resolved. The second issue of this publication series, published in January 2016, already focused on this topic under the title: „Clinical Studies – Which Endpoints Count?" In his opening presentation at the time, Thomas Kaiser, then Head of Pharmaceuticals Assessment at IQWiG, called for a change in thinking, stating: „It is not what has been investigated that is relevant, but what is relevant that should be investigated."

The term „patient-relevant endpoints" is anchored in sub-statutory regulation and was first introduced at the end of 2010 in the Pharmaceutical Benefit Assessment Regulation, which was designed to complement AMNOG. Interestingly, Section 2 of this regulation, titled „Definitions", does not contain a definition of patient-relevant endpoints. This is only implicitly introduced in paragraph 5, which outlines the concept of additional benefit, on the one hand in relation to the categories of mortality, morbidity and quality of life (Section 5, Paragraph 2) and on the other hand in relation to the respective quantification and classification of additional benefit, such as cure, prolongation of survival, freedom from symptoms, or avoidance of side effects (Section 5, Paragraph 7).

There is broad agreement among all stakeholders that the endpoints of clinical studies – and thus the endpoints underlying AMNOG procedures – should be relevant for the affected patients. Furthermore, there is consensus across all parties that many commonly used endpoints, such as overall survival, pain, and severe side effects should be considered relevant for patients.

However, disagreements persist, primarily in the context of the discussion around surrogate parameters and surrogate endpoints. These terms are neither mentioned nor defined in the AMNOG law text or the German Pharmaceutical Benefit Assessment Regulation. The discussion on surrogates only developed in the context of the first AMNOG procedures and was subsequently incorporated into methodological guidelines, such as those of IQWiG. The focus is primarily on endpoints that reliably and directly reflect concrete changes in health status. Surrogate endpoints can be classified as substitutes that allow earlier and simpler representation of the actual patient-relevant endpoints. However, as the informative value of surrogate endpoints is not always reliable, they are only used in the AMNOG after successful statistical validation.

It remains unclear – at least on the part of the regulators – how it can be determined and who has the authority to define certain endpoints as primarily patient-relevant or as surrogates when measuring additional benefits or side effects. Examples in the field of oncology are progression-free survival or the response rate, in metabolic diseases laboratory values such as HbA1C or enzyme activities, in infections the SVR (Sustained Virologic Response), in renal diseases the eGFR or creatinine clearance.

The recent meeting of the Platform for Benefit Assessment and this publication cover the entire spectrum of the associated discussion from different perspectives. In my view, three key insights emerge from reading the articles:

- The debate on the relevance of endpoints for patients is important and will need to continue – both in the context of the national and the developing European HTA process.
- It is neither expected nor necessarily desirable for the underlying complex issues were to be decided on a categorical political or legal basis. This means that interdi-

sciplinary dialogue involving experts such as patients, clinicians, regulators, HTA institutions, methodologists and industry representatives is essential to ensure the quality and subsequent broad acceptance of the assessment both nationally and at European level.

- The classification of certain outcomes as patient-relevant has far-reaching implications for clinical care, extending well beyond AMNOG procedures. This is not just a matter of detailed technical discussions on the applicability of certain statistical-methodological tools in the context of surrogate validations, but also a clinical-qualitative, patient-centred discussion on what constitutes a relevant benefit for patients in research, approval, HTA assessment and clinical care.

My sincere thanks go to the speakers and authors of this publication, who wrote their articles „on top" of their other duties showing their great commitment that cannot be overestimated.

And to you, dear readers, I hope you enjoy and gain valuable insights from these articles.

Jörg Ruof

Contakt: joerg.ruof@r-connect.org

Benefit assessment – which outcome measures are relevant for patients?

Stefan Schwartze, MdB | Federal Government Commissioner for Patients' Affairs

As the Federal Government Commissioner for Patients' Affairs, my goal is to improve the quality of medical treatment and make healthcare more patient-centred so that our healthcare system better meets the needs and requirements of patients.

To achieve this, it is both right and necessary to assign high importance to patient-relevant endpoints in the evaluation of medical services – alongside the available evidence, which holds the highest priority.

But what exactly constitutes a patient-relevant additional benefit, and how is it determined and agreed upon? How are differences in assessment among various decision-makers handled?

Relevant endpoints naturally include the improvement of health status, the reduction of disease duration, the extension of survival, the reduction of side effects, or an improvement in quality of life.

But what about surrogate endpoints – those substitute endpoints such as blood values? There are differing evaluations and assessments among decision-makers, including IQWiG, G-BA, and EMA. More than ten years after the introduction of AMNOG, this has not changed. It would be desirable and timely to establish a procedure that ultimately leads to a binding agreement for all.

But it is also important to me to capture the experiences and perspectives of patients more comprehensively. In addition to other data sources, instruments for the systematic collection of patient-reported treatment outcomes should also be used. After all, it is the patients who experience the treatment process from the very beginning.

Their experiences regarding their treatment and whether they perceive their therapy as effective and successful are rarely asked about – let alone systematically analysed and used to improve healthcare or research.

Regularly collecting patient perspectives on the course of therapy and their experiences with treatment outcomes could make an important contribution to enhancing the effectiveness of medical interventions and improving adherence.

In the long term, I hope that, in addition to patient-relevant endpoints, the experiences and perspectives of patients will also be taken into account and incorporated into assessments.

Stefan Schwartze, *Industrial mechanic with a technical diploma and many years of professional experience; member of the SPD since 1994; chairman of the SPD in the Herford district since 2006; chairman of the SPD region of East Westphalia since 2014. From 1999 to 2009, he served as a district councillor; from 2004 to 2009, he was deputy chairman of the SPD district council faction. Since October 2009, he has been a member of the German Bundestag (MdB). From 2009 to 2022, he was a member of the Petitions Committee; from 2013 to 2022, he was the spokesperson for the SPD parliamentary group's Petitions Working Group. Since 12 January 2022, he has served as the Federal Government Commissioner for Patients' Affairs.*

Patient-relevant endpoints: Key questions from the patient's perspective

Dr Martin Danner, BAG Selbsthilfe | Eva Stumpe, German Society for Muscular Dystrophy (DGM)

One of the most significant advancements brought about by AMNOG was the evaluation of innovations in the pharmaceutical sector based on whether they demonstrate an additional benefit relevant to patients. The emphasis here is on patient relevance. Not just any additional benefit, but only one that is noticeable to patients is considered relevant.

The AMNOG has thus clearly placed the patient perspective at the centre of considerations. On the other hand, other endpoints, particularly the so-called surrogate parameters, have been downgraded in importance.

However, this prioritization of endpoints repeatedly gives rise to discussions and leads to controversies. For example, there is ongoing debate about whether viral load in hepatitis or tumour growth are, in themselves, patient-relevant issues, even though they do not necessarily result in any noticeable „experience" in the everyday sense. Other aspects of patient relevance seem to lead a permanent shadowy existence, particularly the so-called Patient Reported Outcomes (PRO).

In the context of introducing a European benefit assessment procedure for pharmaceuticals, the discussion about the relevance of specific endpoints is becoming even more significant. It is foreseeable that the different discourses in the European countries on this matter will collide and must therefore be reorganised.

What will therefore be relevant in the future to demonstrate the difference in benefit? How will the discussion on the patient relevance of endpoints evolve in Germany? The situation of long-term observations that characterises ATMPs (Advanced Therapy Medicinal Products) raises new questions about the patient relevance of endpoints.

From the patient's perspective, the following questions are currently particularly pertinent:

1) According to the IQWiG's methods paper, the patient relevance of an endpoint is clearly defined from the patient's perspective. Unfortunately, however, it is only rarely successful to make so-called PROs the focus of benefit comparisons. Why is there often a lack of validated measurement tools, and how can progress be made in this area?

2) Recently, the patient relevance of endpoints has been methodologically questioned in part because certain effect differences are not considered sufficient. This is even discussed in the case of natural scales. According to this, a discussion of body size in growth disorders should not be regarded as patient-relevant per se. Is this adequate?

3) Why are differences in the form of administration not considered patient-relevant, even though they can significantly impact patients' lives?

The following additional aspects aim to deepen these questions:

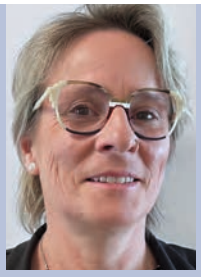
4) What role do real-world data (RWD), which represent the lived perspective of patients outside clinical settings, play in the benefit assessments and the decisions as to whether an endpoint is relevant to patients? Especially in rare diseases, these data, when already available during the benefit assessment, should complement data collected in randomised clinical trials to provide a more valid picture of the actual additional benefit in a broader patient community.

5) If RWD are used, how can it be ensured that patient representatives and patient organisations are involved in the creation and management of the registries in which RWD are ideally collected? Only then can it be guaranteed that these registries include patient-relevant data.

6) Are the validated questionnaires used to determine treatment and health-related quality of life really patient-centred (to what extent did patient organisations or patients participate in their creation)?

7) The validation of a Patient Reported Outcome Measure (PROM) is complex and very costly. In rare diseases, it is often necessary to use validated PROMs that were originally developed for a different condition and are not fully applicable to the rare disease. Shouldn't the healthcare system be responsible for financially and logistically supporting patient organisations in such cases, enabling them (possibly in cooperation with other stakeholders) to develop and validate new PROMs that are better tailored to their specific disease area?
8) Already in the early consultations between the G-BA and pharmaceutical company, the data that should ideally be included in the dossier are discussed. Both IQWiG and the G-BA emphasize the patient relevance of endpoints, and the data provided. However, this patient relevance should then also be requested directly from the respective patient community. In reality, the decisions on patient relevance are made in committees and panels that include only a small proportion of patient representatives. No disease-related patient representatives are involved in the early consultation of the G-BA with the pharmaceutical company.

Dr Martin Danner *is a lawyer and the national managing director of the Federal Association of Self-Help for People with Disabilities and Chronic Illnesses and their Relatives (BAG SELBSTHILFE). After his studies in Heidelberg, he worked as a lawyer for several years specialising in health law before taking over as head of the health policy and self-help promotion department of BAG SELBSTHILFE. He is the spokesman for patient representation at the G-BA and, among other things, participates in the Scientific Advisory Board of the Medical Centre for Quality in Medicine (AZQ) and in the IQWiG Board of Trustees.*



Eva Stumpe *is a patient representative in the field of spinal muscular atrophy. She is a contact person at the DGM e.V., a member of the management team of the SMA Initiative as well as treasurer and board member of SMA Europe e.V. Mrs Stumpe is the mother of two adult children. Her daughter Sarah lives with spinal muscular atrophy type II. For two years, she represented the interests of SMA patients in approval procedures at the EMA as a patient representative. She is now a patient representative on several industry advisory boards. Professionally, she is a lawyer and is involved in the family business.*

Primary endpoints in the AMNOG: status quo and outlook using oncology as an example

Professor Bernhard Wörmann | German Society for Haematology and Oncology (DGHO)

The definition of relevant endpoints is a critical component of the design of clinical studies for new pharmaceuticals in terms of content and methodology. The selection and prioritisation of endpoints determine the study design, the measurement parameters to be collected, the required number of study participants, discontinuation criteria, and ultimately, the access of patients to a new pharmaceutical. The study results are used both in approval and benefit assessment procedures as well as for recommendations in guidelines – but they are not evaluated using the same methodology. Over the last 13 years, extensive experience has been gained from the AMNOG procedures in Germany, with more than 300 completed procedures on new pharmaceuticals or new indications in oncology alone. This article presents and discusses the strengths and weaknesses of the most common primary endpoints.

Introduction

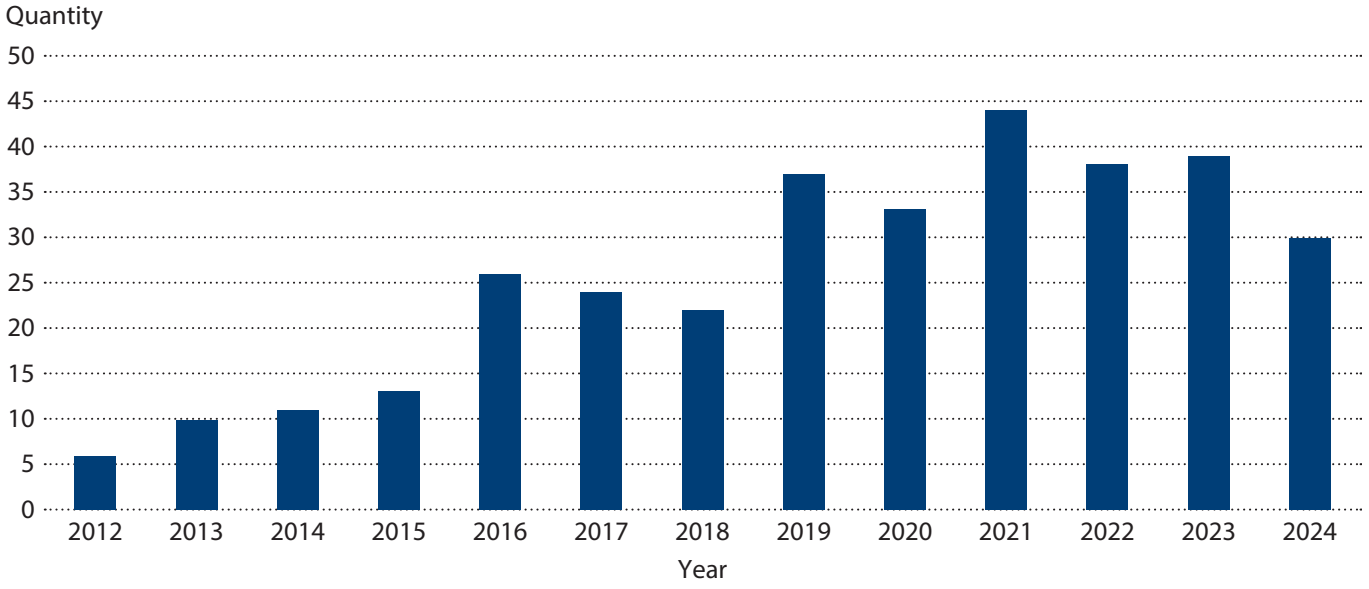
In the past two decades, drug-based therapy has significantly improved the prognosis of patients with cancer. The ultimate goal in the treatment of these patients is a cure without negative consequences. The prerequisites are the elimination of the cause of the disease and the prevention of recurrence using therapy strategies with few or no side effects. Today, this goal is achievable for many cancers. The chances of a long-term cure have been increased, particularly in patients with localised disease and a high risk of recurrence, by combining effective local measures such as surgery and/or radiotherapy with the use of systemic therapy. Prime examples include breast, lung, and colorectal carcinoma.[1-6]

The relatively greatest progress has been achieved in patients with advanced, metastatic, or primarily systemic cancers. A new experience in the use of highly effective pharmaceuticals in this patient group is the increasing transformation of very advanced cancers, which previously had a life expectancy limited to months or a few years, into chronic diseases with an almost normal life expectancy. Prime examples include metastatic melanoma[7], biologically defined subgroups of metastatic non-small cell lung cancer[3,8,9], and haematological neoplasms such as chronic myeloid leukaemia[10], multiple myeloma[11], and chronic lymphocytic leukaemia[12].

Experience from the AMNOG procedure

Between 2012 and 2024, more than 300 procedures on new oncology pharmaceuticals or new therapeutic areas were completed by the Federal Joint Committee (G-BA) as part of the AMNOG process. The total number of completed procedures in oncology exceeds 400, including reassessments, for example after expiry of the deadline.13 Figure 1 provides an overview of the number of completed procedures within

Completed procedures of early benefit assessment in oncology 2012-2024

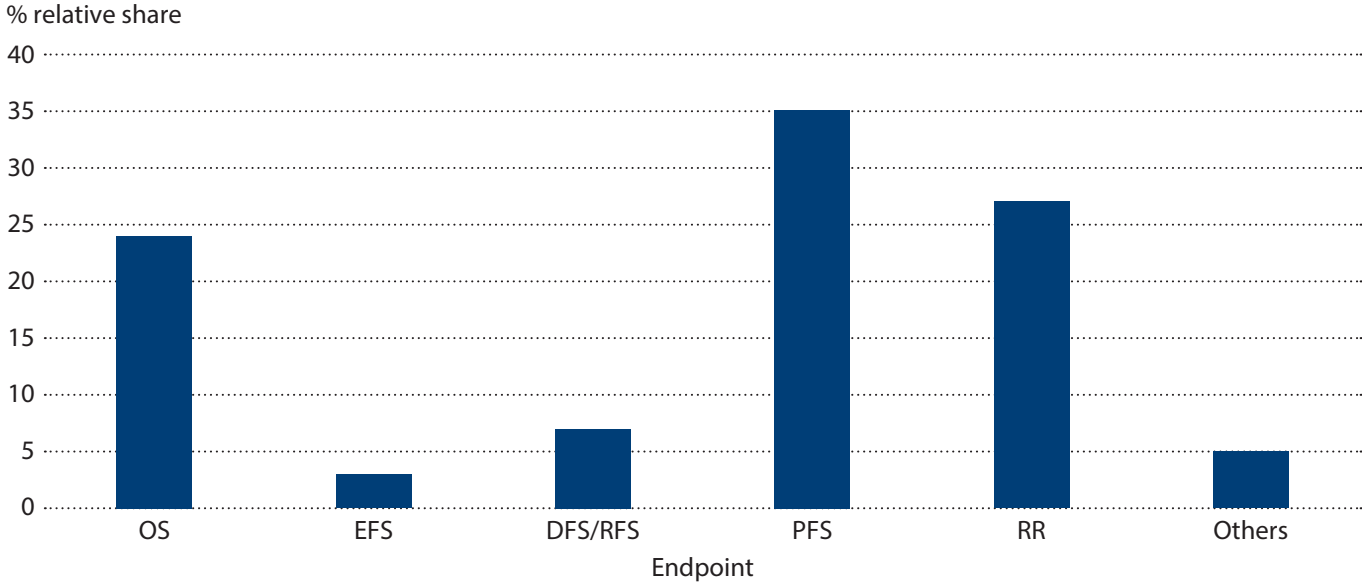


Source: Own presentation

Figure 1: Overview of the number of completed procedures within the framework of early benefit assessments from 2012 to 2024, limited to new pharmaceuticals or new indications.

Professor Bernhard Wörmann *works as a physician specialising in internal medicine, haematology, and internal oncology with additional qualification in palliative care. Since 2010, he has served as Medical Director of the DGHO – German Society for Haematology and Medical Oncology and, since 2015, as Chair of the „Pharmaceutical Benefit Assessment" Commission of the AWMF. He works as a physician at the Charité Outpatient Health Centre and in the Medical Clinic specialising in haematology, oncology, and tumour immunology at the Virchow Campus in Berlin.*

Distribution of primary study endpoints in the completed procedures of the early benefit assessment in oncology 2012-2024



OS: overall survival; EFS: event-free survival; DFS/RFS: disease/relapse-free survival; PFS: progression-free survival; RR: response rate; Others: e. g. toxicity, quality of life

Source: Own presentation

Figure 2: In the majority of the 333 evaluated AMNOG procedures, a single primary endpoint was defined for the respective study. Two co-primary endpoints were used in 51 procedures.

the framework of early benefit assessments from 2012 to 2024, limited to new pharmaceuticals or new indications.

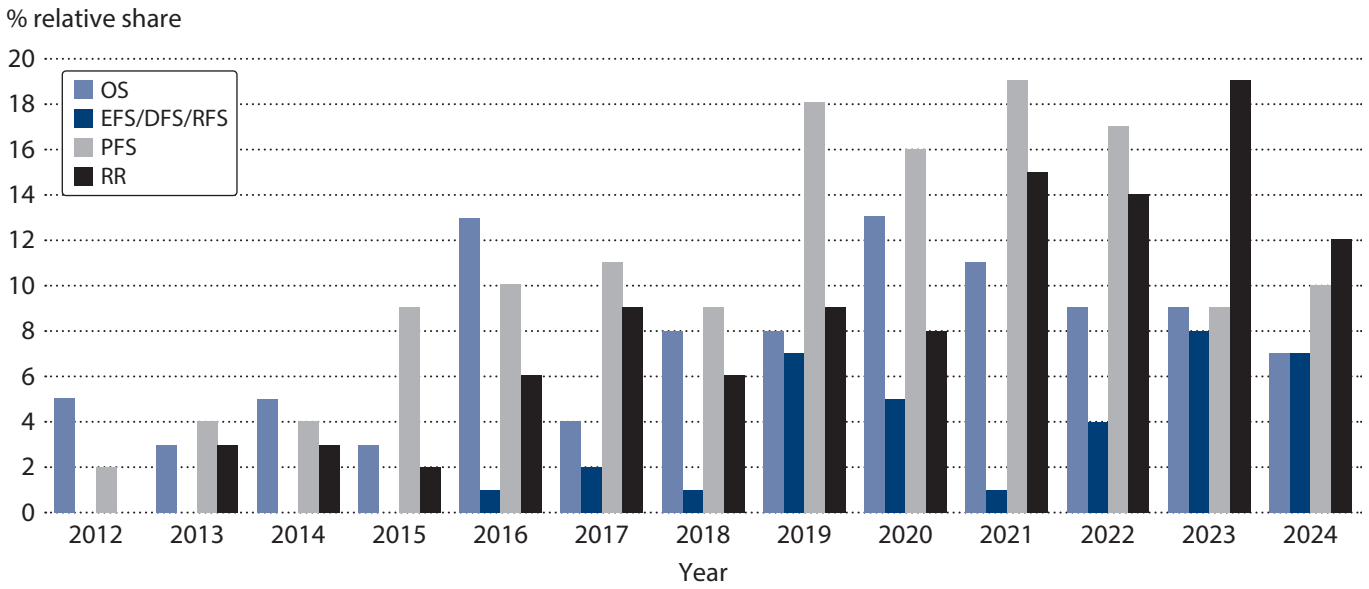
Primary endpoints

In oncology, numerous different endpoints are used in approval studies. The definition of the primary endpoint is of particular importance. In the vast majority of the 333 evaluated procedures, a single primary endpoint was defined for the respective study. In 51 procedures, two co-primary endpoints were used, for example overall survival and progression-free survival. Figure 2 provides an overview of the number of endpoints used in completed procedures between 2012 and 2024.

In the period analysed, there was a clear shift in the frequency of use of primary endpoints (figure 3).

Descriptively, the number of procedures with the primary endpoint „overall survival (OS)“ has remained quite constant, while the number of procedures with the endpoint „progression-free survival (PFS)“ has significantly increased, especially in the years 2019 to 2022. In recent years, the number of procedures with the primary endpoint „response rate (RR)“ has also risen.

Distribution of primary endpoints in the completed procedures of the early benefit assessment in oncology 2012-2024



OS: overall survival; EFS/DFS/RFS: event-free/disease-free/relapse-free survival; PFS: progression-free survival; RR: response rate

Source: Own presentation

Figure 3: Although the prolongation of overall survival is the most crucial endpoint for oncology patients in the vast majority of indications, it was the primary endpoint in only 24% of completed procedures.

Mortality - overall survival time

Prolongation of overall survival is the most crucial endpoint for oncology patients in the vast majority of indications. However, it was the primary endpoint in only 24% of completed procedures. The distribution of this endpoint has not changed significantly during the observation period (figure 3). The overall survival time is relevant for the assessment of the efficacy of a new pharmaceutical in patients, but in the current standard form for a benefit assessment it is fraught with weaknesses.

Strengths:

- Objective measurement
- Good comparability with data from other studies or registries
- High patient relevance

Weaknesses:

- Survival rate: The survival rate is rarely evaluated as a primary endpoint or at all. The survival rate is particularly relevant for the evaluation of an oncological study if there is a chance of a sustained extension of overall survival. This has been particularly relevant in recent years

for studies with immune checkpoint inhibitors.

- Feasibility of studies: A very high number of study participants is required to detect a significant difference, especially for diseases with a good prognosis.
- Potential for bias: Measuring overall survival requires capturing all subsequent therapies within the indication. In view of the quite different access to other new pharmaceuticals worldwide, there is a high potential for bias. Moreover, in some studies concerning end-of-life situations, a crossover design is ethically required. There is currently no generally accepted consensus on the best methodology for calculating crossover effects. With high crossover rates, the content of the research question changes: the value of a new pharmaceutical or therapy concept is no longer evaluated on its own but rather in comparison to its use in an early versus a late disease stage.
- Lack of validation: In many HTA methodologies, the extent of the relative difference in median survival time is the relevant assessment parameter.[14] Thus, even in 2024, an absolute difference of <3 months led to the proposal of a significant added benefit, without a sustained increase in survival time and, above all, without an increase in the long-term survival rate.[15]

Morbidity - disease/relapse-free survival (DFS/RFS) / event-free survival (EFS)

Disease-free survival (DFS/RFS) and event-free survival (EFS) are useful and commonly used endpoints for all therapeutic interventions with a curative intent. These endpoints are used for systemic diseases, such as acute leukaemia as well as for neoadjuvant and adjuvant therapy.

Strengths:

- Direct mapping of the effectiveness of the respective intervention
- Shorter time interval to study endpoint compared to overall survival time
- Less influence from later confounders such as relapse therapy or comorbidities affecting overall survival.

Weaknesses:

- Composite endpoint: All composite endpoints are associated with a higher risk of bias due to different weighting of included parameters. In a system of benefit assessment such as in Germany, where HTA assessment and pricing are essentially based on comparison with the previous standard of care, standardisation of the definition of composite endpoints is essential, but also more difficult to achieve.
- Overlapping of the detection of the efficacy of a new pharmaceutical by high treatment-related toxicity and lethality.

Morbidity - Progression-free survival (PFS)

Progression-free survival (PFS) or time to progression has been the most frequently used primary endpoint in recent years for assessing the efficacy of new oncology pharmaceuticals (figure 1). PFS is a meaningful and common endpoint for therapeutic interventions with non-curative intent. In recent years, the European Medicines Agency (EMA) and the US Food and Drug Administration (FDA) have accepted PFS as an approval criterion for various oncology pharmaceuticals and therapy indications. Regulatory and HTA bodies have extensively examined its methodological principles.

Strengths:

- Direct mapping of the effectiveness of the respective intervention in advanced diseases and non-curative therapies
- Rapid assessment of efficacy
- Reduction in required study participants compared to the overall survival endpoint.

Weaknesses:

- Composite endpoint: All composite endpoints are associated with a higher risk of bias due to different weighting of included parameters.
- Need for transparent presentation of censoring; The estimated treatment effect increases with the degree of censoring, which may suggest a greater treatment effect than actually exists.[17]
- Reliance on technical endpoints such as imaging poses the risk of distortion by other therapy-induced effects, such as pseudo-regression and pseudo-progression[20] [de Groot], and/or by laboratory results, which may be influenced by therapy-related side effects.
- Very heterogeneous picture in the correlation of PFS with overall survival in different tumour entities.[18]

The objection of a lack of patient relevance of the primary endpoint PFS can be mitigated by combining it with the evaluation of secondary endpoints such as symptoms, quality of life (patient-reported outcomes) and toxicity.

Morbidity - response rate

In aggressive malignancies, an extension of the survival is only possible, if remission has been achieved, usually complete remission. This can be achieved in metastatic or primarily systemic diseases through pharmacotherapy. In case of indolent malignancies, this correlation is less clear. The assessment of the response rate as a primary endpoint per se is difficult because its use in the context of approval studies is primarily related to the quality of the studies (figure 4).

The graph illustrates that response rate was mainly used as a primary endpoint in non-randomised studies, the number of which has increased in recent years.

Strengths:

- Direct mapping of the effectiveness of the respective intervention
- Rapid assessment of efficacy
- Reduction in the number of study participants required for the endpoints overall survival, disease-free survival, and progression-free survival.

Weaknesses:

- Reliance on technical endpoints such as imaging poses the risk of distortion by other therapy-induced effects, such as pseudo-regression and pseudo-progression [de Groot], and/or by laboratory results, which may be influenced by therapy-related side effects.
- No direct correlation with patient-reported outcomes.
- Weak correlation with overall survival.

Morbidity – symptoms

In oncology, there is a wide range of highly distressing symptoms that significantly impact quality of life. Pharmaceuticals that provide effective symptom relief or significantly extend the time until the onset of distressing symptoms are of great value. Nevertheless, symptoms are rarely defined as a primary endpoint, specifically in only 28 of the 333 evaluated procedures. In Figure 1, these studies are grouped under „other". One of these exceptional examples is the stool frequency in patients with advanced carcinoid tumours when assessing telotristat ethyl, as well as symptoms in patients with primary myelofibrosis due to fedratinib, momelotinib, or ruxolitinib.

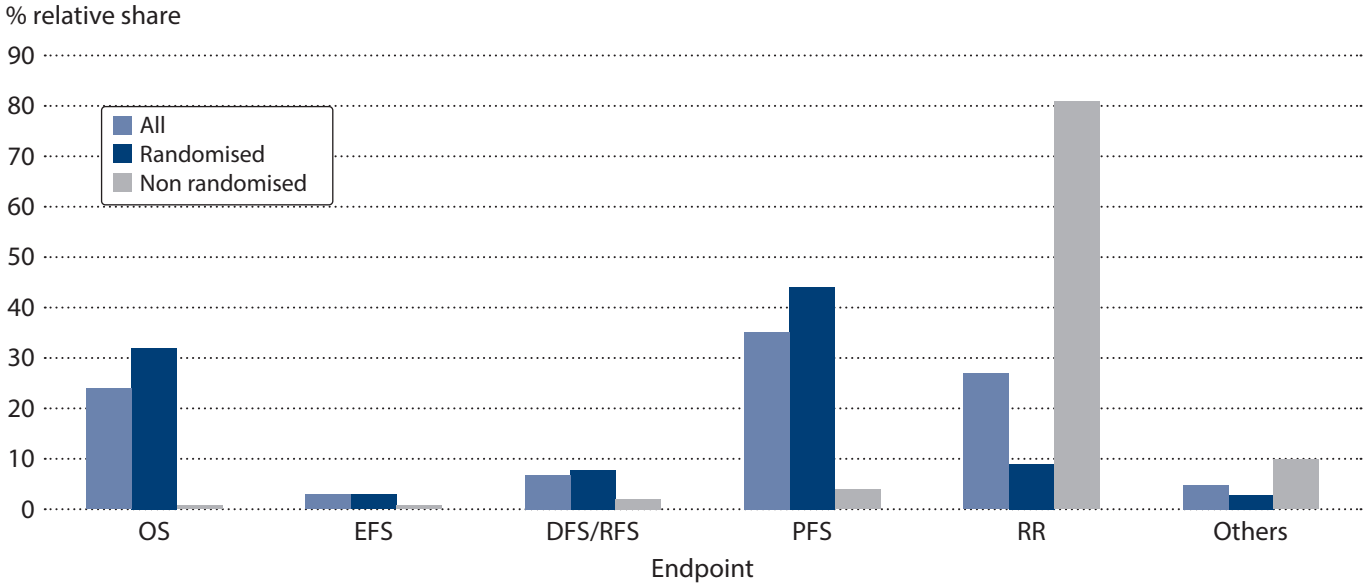
Quality of life / Patient-Reported Outcome (PRO)

Quality of life parameters were not the primary or co-primary endpoint in any of the oncological studies.

Summary and outlook

Oncology is a good example of the necessary diversity of endpoints required when assessing and evaluating the benefits of a new pharmaceutical. Figure 5 provides an

Correlation between study design and selection of primary endpoints



OS: overall survival; **EFS:** event-free survival; **DFS/RFS:** disease/relapse-free survival; **PFS:** progression-free survival; **RR:** response rate; Others: e. g. toxicity, quality of life

Source: Own presentation

Figure 4: The response rate was mainly used as a primary endpoint in non-randomised studies. Their number has increased in recent years.

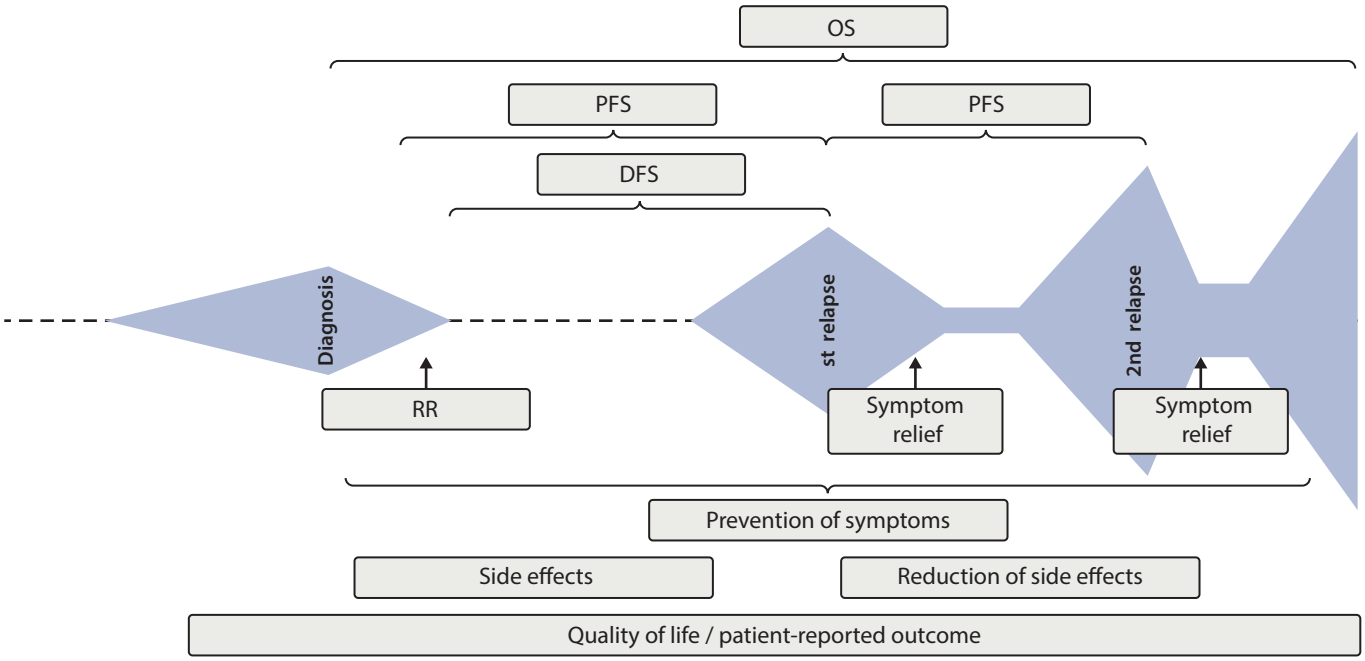
overview of patient-relevant endpoints based on the course of the disease.

New pharmaceuticals are assessed as the basis for marketing authorisation, benefit assessment as the basis for pricing and in guidelines. Each procedure has its own methodology. From a healthcare perspective, it is important to avoid „parallel worlds" with absolute claims in pharmaceutical evaluation. Contradictory or incomprehensible assessments create uncertainty among patients and prescribing physicians.

Furthermore, it is crucial to assess the value of endpoints. Based on experience from over 300 procedures concerning new pharmaceuticals and indications, the following challenges arise for HTA evaluations within the AMNOG process in oncology:

- Adapting HTA criteria to longer survival times, for example through landmark analyses and/or greater integration of additional patient-relevant endpoints
- Developing and validating a methodology for weighted evaluation of different endpoints, with direct patient involvement
- Review of the criteria for non-inferiority studies.

Patient-relevant endpoints in oncology



OS: overall survival; **DFS:** disease-free survival; **PFS:** progression-free survival; **RR:** response rate

Source: Own presentation

Figure 5: Oncology, here based on disease progression, serves as an example of the required diversity of endpoints for assessing and evaluating the benefits of a new pharmaceutical.

References

[1] Working Group on Gynaecological Oncology: Diagnostik und Therapie primärer und metastasierter Mammakarzinome: Endokrine und zielgerichtete Therapie metastasiertes Mammakarzinom. Status March 2024. (Diagnostics and therapy of primary and metastatic breast cancer: Endocrine and targeted therapy for metastatic breast cancer.). AGO_2024D_26_Therapiealgorithmen_MASTER_schwarz (ago-online.de)

[2] AWMF-Guideline: Früherkennung, Diagnostik, Therapie und Nachsorge des Mammakarzinoms. Registry number 032-045OL, July 2021. (Early detection, diagnosis, treatment, and follow-up of breast cancer). https://go.sn.pub/flkeby

[3] Griesinger F et al.: Nicht-kleinzelliges Lungenkarzinom (NSCLC). (Non-Small Cell Lung Cancer (NSCLC). Guidelines by DGHO, OeGHO, SGMO, and SGH+SSH, Status January 2025. https://www.dgho-onkopedia.de/de/onkopedia/leitlinien/lungenkarzinom-nicht-kleinzellig-nsclc

[4] Interdisziplinary S3 guideline: Prävention, Diagnostik, Therapie und Nachsorge des Lungenkarzinoms, 020-007, November 2022. (Interdisciplinary S3 Guideline: Prevention, Diagnosis, Therapy, and Follow-Up of Lung Cancer). https://go.sn.pub/nzyfwn

[5] Colorectal Cancer. Guidelines by DGHO, OeGHO, SGMO, and SGH+SSH, Status January 2024. https://go.sn.pub/mdg4um

[6] https://go.sn.pub/i1v82z

[7] Wolchok JD, Chiarion-Sileni V, Rutkowski P et al.: Final, 10-Year Outcomes with Nivolumab plus Ipilimumab in Advanced Melanoma. N Engl J Med 392:11-22, 2025. DOI: 10.1056/NEJMoa2407417

[8] Wu YL, Dziadziuszko R, Ahn JS et al.: Alectinib in Resected ALK-Positive Non-Small-Cell Lung Cancer: N Engl J Med 390_1265-1276, 2024. DOI: 10.1056/NEJMoa2310532

[9] Planchard D, Jänne PA, Cheng Y et al.: Osimertinib with or without Chemotherapy in EGFR-Mutated Advanced NSCLC. N Engl J Med 389:1935-1948, 2023. DOI: 10.1056/NEJMoa2306434

[10] Hochhaus A, Larson RA, Guilhot F, et al. Long-term outcomes of imatinib treatment for chronic myeloid leukemia. N Engl J Med 376:917-927, 2017. DOI: 10.1056/NEJMoa1609324

[11] Sonneveld P, Dimopoulos MA, Boccadoro M et al.: Daratumumab, Bortezomib, Lenalidomide, and Dexamethasone for Multiple Myeloma. N Engl J Med 390:301-313, 2024. DOI: 10.1056/NEJMoa2312054

[12] Wendtner C et al.: Chronische lymphatische Leukämie, September 2024. (Chronic Lymphocytic Leukaemia). Onkopedia – Chronische Lymphatische Leukämie (CLL) (Chronic Lymphocytic Leukaemia).

[13] Benefit Assessment under § 35a SGB V (filtered) – Federal Joint Committee (G-BA)

[14] General Methods - Version 7.0

[15] Benefit assessment procedure for the active substance fruquintinib (colorectal carcinoma, previously treated patients) - Federal Joint Committee

[16] European Medicines Agency: Guideline on the clinical evaluation of anticancer medicinal products, 18 November 2023. Anticancer GL Rev 6 adopted CHMP November 2023

[17] Lengliné E, Baba J, de Boissieu P et al.: Composite event-free-survival as an endpoint in oncology drug evaluation: Review and guidance perspectives from the Haute Autorité de Santé (HAS). Eur J Cancer 204:114047, 2024. DOI: 10.1016/j.ejca.2024.114047

[18] Belin L, Tan A, de Rycke Y, Dechartres A: Progression-free survival as a surrogate for overall survival in oncology trials: a methodological systematic review. Br J Cancer 122:1707-1714, 2020. DOI: 10.1038/s41416-020-0805-y

Patient relevance: a legal perspective

Professor Sebastian Kluckert | Bergische Universität Wuppertal – Chair of Public Law, in particular Public Commercial Law and Social Law | Impartial Chairman of the Arbitration Board pursuant to Section 130b (5) SGB V

The term „patient relevance" plays a central role in the context of the benefit assessment of pharmaceuticals and is subject to the discretionary judgement of the Federal Joint Committee (G-BA). Courts review the G-BA's assessment only for its reasonableness. The limits of reasonableness are exceeded if decisions violate recognised standards of evidence-based medicine and health economics. Diverging opinions within specialist disciplines increase the G-BA's scope for judgement. The debate on the recognition of certain endpoints is, therefore, not primarily a legal issue.

Patient relevance" is a legal term derived from the law governing the benefit assessment of pharmaceuticals. This legal framework is based on parliamentary legislation, specifically Section 35a SGB V, and is further specified by subordinate regulations and procedural rules of the Federal Joint Committee (G-BA). While the marketing authorisation of pharmaceuticals primarily concerns patient safety and efficacy, the benefit assessment of pharmaceuticals serves as the basis for pricing new active substances (cf. Section 130b SGB V, Section 78 (3a) AMG). The pharmaceutical price or reimbursement amount should be commensurate with the benefit in relation to an appropriate comparative therapy.

The legal term patient relevance is found in the following regulations, among others: According to Section 2 (3) of the Regulation on the Benefit Assessment of Medicinal Products under Section 35a (1) SGB V for Reimbursement Agreements under Section 130b SGB V (AM-NutzenV), „the benefit of a medicinal product […] is the patient-relevant therapeutic effect, particularly regarding the improvement of health status, reduction in disease duration, extension of survival, reduction of side effects, or improvement of quality of life". Section 5 (5) sentence 1 AM-NutzenV states that „the additional benefit compared to the appropriate comparator therapy is determined as an improvement in the influence on patient-relevant endpoints for benefit in accordance with Section 2 (3)" AM-NutzenV. According to Section 35a (3b) sentence 4 SGB V, the G-BA must „define requirements for patient-relevant endpoints and their collection" as part of application-accompanying data collection and evaluation.

Scope of judgement and reasonableness review

In the application of law, the key question is whether a

specific, individual factual circumstance (e.g. an improvement in progression-free survival demonstrated by studies) falls under an abstract, general legal term (e.g. „patient-relevant endpoint"). Lawyers refer to this as determining whether the factual circumstance can be subsumed under the law. This essentially means assessing whether the legislator intended for the legal norm to regulate the given factual circumstance. Ultimately, courts have the final authority in answering this question.

However, the courts' competence is limited when the law – albeit rarely – grants an administrative body discretionary power, allowing it to determine, in preference to the courts, whether a factual circumstance and the legal provision align. This room for manoeuvre is known as scope for judgement. Where such a scope for judgement is recognised, courts examine whether the authority has complied with jurisdictional, procedural, and formal requirements, relied on the correct facts, and refrained from extraneous considerations. Otherwise, however, there is only a review of reasonableness.

„Patient relevance" is a so-called undefined legal term. Undetermined legal terms are legal terms that are highly open to interpretation, are difficult to define due to their lack of specificity and are based on complex judgements, evaluations, and considerations. Examples include terms such as „suitability", „special case of hardship", or „defacement of the landscape". Unlike purely descriptive legal terms (such as „dog", „cat", „mouse"), it is not immediately clear whether a given factual circumstance falls under an indeterminate legal term. Undefined legal terms provide a potential gateway for recognising a scope for judgement granted to an authority by the legislator.

In the context of the benefit assessment of pharmaceuticals, the Federal Social Court (Bundessozialgericht, BSG) acknowledges the scope for judgement. Benefit assessment resolutions are classified by the BSG as acts of subordinate legislation. This follows from the legal nature of the G-BA's Pharmaceutical Directive (Section 92 (1) sentence 2 no. 6 SGB V), of which benefit assessment resolutions form a part (Section 35a (3) sentence 7 SGB V). The scope for judgement is derived from the legislator's authority to shape regulations. Consequently, judicial review in social court proceedings may not replace the G-BA's assessment with its own (cf. BSG, judgement of 12 August 2021 – B 3 KR 3/20 R = BSGE 133, 1 para. 33). However, this limitation applies only as long as the G-BA's assessment remains reasonable (reasonableness review).

Professor Sebastian Kluckert *studied business administration and law in Berlin. After completing his doctorate and legal clerkship, he worked as a research assistant at the Department of Law at the Free University of Berlin. In December 2016, he completed his habilitation with authorisation to teach constitutional and administrative law, European law, public commercial law, and social law. He has been a professor at the University of Wuppertal since April 2018. Previously, he also worked as a lawyer in a law firm specialising in healthcare. He has been the impartial chairman of the AMNOG Arbitration Board since 1 July 2023.*

Limits of reasonableness

Where do the limits of reasonableness lie when determining whether a specific endpoint falls under the legal term „patient relevance"? An assessment would be unreasonable if it classified an endpoint recognised in higher-ranking law (particularly SGB V, AM-NutzenV) as not patient-relevant. Sections 2 (3) and 5 (2) sentence 3 AM-NutzenV contain exemplary regulations – explicit but non-exhaustive factual situations that, in the legislator's view, generally fall under the legal term. In this case, these are the endpoints of health status/morbidity, duration of illness/morbidity, survival/mortality, side effects and quality of life. This means that whether a disputed endpoint is relevant in the context of benefit assessment can also be politically determined by the legislator or regulatory authority.

If no contradiction with higher-ranking law exists, is it sufficient for reasonableness that a given endpoint is deemed non-patient-relevant within the relevant scientific disciplines (particularly medicine)? If this question were answered affirmatively, the G-BA could justify its assessment based on credible expert opinions, even if they contradict the predominant scientific consensus. According to the prevailing view, a G-BA assessment that contradicts the generally accepted standards of health technology assessment (HTA) science is not reasonable. This view can be derived from Section 35a para. 1 sentence 8 no. 2 SGB V and Section 7 para. 2 AM-NutzenV.

According to these provisions, the international standards of evidence-based medicine and health economics form the basis of the benefit assessment. The benchmark for the assessment as part of the benefit assessment is the generally recognised state of medical knowledge. Therefore, in addition to the statutory standard examples, endpoints that are recognised as patient-relevant according to international standards must also be recognised as patient-relevant by the G-BA. In my opinion, this also includes endpoints that are recognised as meaningful surrogates of a patient-relevant endpoint (mentioned in the law or other internationally recognised endpoints) according to international standards.

With regard to surrogates, however, earlier decisions of the BSG and the Berlin-Brandenburg Regional Social Court (Landessozialgericht Berlin-Brandenburg), which is responsible for the benefit assessments of the G-BA at first instance (between 2011 and 2013), often state that studies that have formulated mere surrogate parameters as the primary objective cannot be considered from the outset to prove a therapeutic improvement (most recently BSG, judgement of 17 September 2013 - B 1 KR 54/12 R = BSGE 114, 217 para. 48; probably most recently LSG Berlin-Brandenburg, judgement of 7 June 2013 - L 7 KA 164/09 KL, juris para. 126). This categorical exclusion has probably been tacitly abandoned in the meantime.

Possible impetus from Union law

Impulses for the justifiability of the assessment of the patient relevance of an endpoint can also result from the European benefit assessment (EU HTA). Methodologically, a binding impulse could arise as a result of an interpretation and application of national law in conformity with EU law (AM-NutzenV). Alternatively, such an impulse could also unfold below the strict binding force resulting from the primacy of application of EU law in such a way that the recognition of an endpoint as patient-relevant within the framework of the EU HTA is regarded by national jurisdiction as proof or as a strong indication that the endpoint is patient-relevant – according to the international standards of evidence-based medicine and health economics or according to the recognised state of medical knowledge.

However, both impulses would probably require a

further development of the current EU HTA process. But according to Regulation (EU) 2021/2282 on health technology assessment, it is already not completely ruled out today that the ECJ could grant some kind of EU law effect on national assessment procedures to the definition of the relevant parameters for the scope of assessment (including health-related endpoints) in the context of joint clinical assessments, even though such an effect is currently not politically intended according to the official categorisation of the scope of Regulation (EU) 2021/2282.

Clarification of patient relevance in legal proceedings

In principle, the relevance of an endpoint to patients can be clarified by way of legal proceedings. Benefit assessments of the G-BA that wrongly (however: scope for judgement!) do not recognise a patient-relevant endpoint for which an improvement occurs according to the study situation are unlawful. A pharmaceutical company can challenge an adverse benefit assessment through legal action (jurisdiction: Berlin-Brandenburg Social Court and Federal Social Court). This challenge can be combined with an appeal against an arbitration ruling on the reimbursement amount under Section 130b (4) SGB V, requesting the annulment of the underlying benefit assessment resolution.

Alternatively, a declaratory action against the G-BA's benefit assessment decision alone is possible if an amicable reimbursement agreement has been reached between the pharmaceutical company and the National Association of Statutory Health Insurance Funds (GKV-Spitzenverband), despite disagreements over the benefit assessment. To date, there have been no known court cases explicitly aimed at establishing the patient relevance of an endpoint

Conclusion

Patient relevance is an indeterminate legal term with a scope for judgement in favour of the G-BA. With the assumption of a margin of judgement and the associated reduced judicial review in the form of a mere justifiability check, case law and legal science throw the ball back into the court of medicine, health economics and other HTA sciences. The dispute as to whether a certain study endpoint is relevant to patients or not is therefore less due to inadequate legal anchoring. Rather, the problem lies in the fact that the qualification of a certain study endpoint as patient-relevant is disputed among the representatives of the relevant HTA sciences. As long as this is the case, the G-BA's refusal to recognise the patient relevance of an endpoint remains justifiable and lawful.

Study endpoints in the early benefit assessment from IQWiG's perspective

Dr Daniela Preukschat, Dr Sebastian Meller | Pharmaceuticals Assessment Department at IQWiG

In early benefit assessment, the central question is whether a new pharmaceutical offers a patient-relevant benefit compared to the current standard of care. Such an advantage can be derived based on patient-relevant endpoints or valid surrogates. The problem is that data to answer this question is still too often lacking. Far too often, there are no suitable studies available for the assessment. If suitable studies do exist, these are usually approval studies, whose primary endpoints are often not patient-relevant or include unvalidated surrogate parameters.

Patient relevance is firmly anchored in early benefit assessment. According to the Ordinance on the Benefit Assessment of Pharmaceuticals (AM-NutzenV),[1] the benefit of a pharmaceutical is defined as „the patient-relevant therapeutic effect, particularly concerning the improvement of health status, the shortening of disease duration, the prolongation of survival, the reduction of side effects, or an improvement in quality of life. The additional benefit, within the meaning of this regulation, is a benefit that is quantitatively or qualitatively greater than the benefit of the appropriate comparator therapy" (Section 2 para. 4, AM-NutzenV).[1] According to IQWiG's methods paper,[2] patient relevance is defined as „how a patient feels, performs his/her functions and activities, or whether he/she survives".

The benefit assessment therefore focuses on endpoints that can be directly perceived by patients. In early benefit assessment, patient-relevant endpoints are categorised into four groups: mortality, morbidity, health-related quality of life, and side effects. In contrast, endpoints based on imaging or laboratory findings, which are not perceptible to patients. Changes to the mode of administration are also repeatedly proposed as a patient-relevant benefit per se. However, without evidence that a new mode of administration influences patient-relevant endpoint such as health-related quality of life, such a benefit cannot be derived either in accordance with SGB V or based on IQWiG's definition of patient relevance. Only in exceptional cases, e.g. if oral administration is possible instead of intrathecal administration,[3] and a reduction in complications due to the administration of a pharmaceutical can be assumed, can a patient-relevant benefit be derived.

High importance of patient-reported outcomes

Patient-reported outcomes (PROs) are of great importance for morbidity and quality of life outcomes, and their role in clinical studies has been strengthened due to their significance in the AMNOG procedure. In PROs, patients themselves assess how the use of the new pharmaceutical affects their symptoms or health-related quality of life. Without the collection of PROs, an incomplete picture of the benefits and risks of new pharmaceuticals is obtained. Patient involvement is essential even during the development of PRO instruments. This is the only way to ensure that all relevant aspects from the patient's perspective are comprehensibly and completely mapped.

PROs should generally be preferred over anthropometric parameters, such as body height as a patient-relevant endpoint. Although body height (z-score) has been classified as patient-relevant in the therapeutic area achondroplasia,[4] it is difficult to assess how a specific change in this endpoint ultimately affects the patient, e.g. on his/her functional limitations and pain. Therefore, the additional benefit in the endpoint of body height (z-score) could not be conclusively quantified in the benefit assessment.

The use of PROs is also possible for endpoints related to side effects but is not yet standard. Moreover, not all side effects are suitable for recording using PROs; for example, directly observable/measurable events (e.g. retinal tear) are unsuitable, whereas subjectively perceived side effects like nausea can be well captured through PROs. For this purpose, the National Cancer Institute (NCI) has developed the PRO-Common Terminology Criteria for Adverse Events (PRO-CTCAE) system for the assessment of symptomatic toxicity in patients in oncological studies.[5] This system has already been presented to IQWiG in several benefit assessments, for the first time in A20-87 and A23-86. However, the corresponding data were not usable due to the lack of detailed justification for the selection of symptomatic AEs

Dr Sebastian Meller, *has been working as a research associate in the Pharmaceutical Assessment Department of IQWiG since 2021. After studying biology at the University of Cologne and completing his doctorate in biomedicine at the University Hospital of Bonn, he worked for several years as a project manager at the Centre for Clinical Trials in Cologne.*

Dr Daniela Preukschat, *has been Head of the Pharmaceutical Assessment Department of IQWiG (Chronic Diseases Division) since 2021. After studying biology and sports science, she completed her doctorate at the Institute of Genetics at the University of Cologne in 2013. She then worked in the evidence-based medicine department of the MDS (Medical Service of the National Association of Statutory Health Insurance Funds, Essen) and joined IQWiG in 2016.*

from the PRO-CTCAE system and because the results were presented only descriptively without considering different observation periods.[6,7]

Only a few surrogate validations submitted

In early benefit assessments, IQWiG has so far accepted three surrogate endpoints as sufficiently valid surrogates for different patient-relevant endpoints: Virological response as a surrogate for AIDS/overall survival in HIV infection, HbA1c level as a surrogate for microvascular complications in type 1 diabetes, and sustained virological response (SVR) as a surrogate for the prevention of the occurrence of hepatocellular carcinoma in hepatitis C infection.[8-10] Overall, however, only a few validations of surrogates were submitted to IQWiG for the early benefit assessment; mainly in oncology (table 1).

Particularly notable is progression-free survival (PFS), for which the most surrogate validations have been submitted

Surrogates accepted by IQWiG and surrogates for which validations were submitted as part of the early benefit assessment

Indication	Endpoints		Surrogate accepted by IQWiG?	Dossier assessment (examples)
	Surrogate	Patient-relevant endpoint		
HIV infection	Virological response	AIDS/overall survival	YES	A12-04 Rilviripine
Diabetes type 1	HbA_{1c}	Microvascular complications	YES	A15-10 Insulin degludec
Hepatitis C	SVR	Hepatocellular carcinoma	YES	A11-17 Boceprevir
Melanoma	PFS	Overall survival	NO	A13-35 Debrafenib, A15-33 Pembrolizumab
Breast cancer	PFS	Overall survival	NO	A16-74 Palbociclib, in A22-66 **no** effect in overall survival (final analysis PALOMA-2)
Breast cancer – adjuvant	DFS	Overall survival	NO	A18-41, A21-11 Pertuzumab, in A22-103 effect in overall survival in 3rd data cut-off
Prostate carcinoma	PFS, MFS	Overall survival	NO	A18-80 Enzalutamid, A20-36 Apalutamid
Interstitial lung disease	Annual forced vital capacity	Overall survival	NO	A20-71 Nintedanib

DFS: disease-free survival; HbA_{1c}: Haemoglobin A_{1c}; **MFS:** metastasis-free survival; **PFS:** progression-free survival; **SVR:** sustained virological response

Source: IQWiG

Table 1: To date, the validations or validation studies presented were largely unsuitable for demonstrating the suitability of the respective surrogate endpoints.

(for melanoma, breast cancer, and prostate cancer[11-15]). Additionally, surrogate validations have been presented for disease-free survival (DFS) in adjuvant breast cancer therapy, metastasis-free survival in prostate cancer, and annual forced vital capacity in interstitial lung disease.[15-18]

However, to date, the validations or validation studies presented were largely unsuitable for demonstrating the suitability of the respective surrogate endpoints for overall survival. Regardless of the question of suitability as a surrogate for overall survival, some outcomes or operationalisations that depict the progression of cancer were classified as patient-relevant: for example, symptomatic progression in prostate cancer, which was operationalised via symptoms noticeable to the patient, was accepted in two benefit assessments.[15,19] Recurrences have also been generally accepted as a patient-relevant endpoints in various benefit assessments.[16,17]

Although the recurrence endpoint is based on imaging, it is considered a failure of the curative approach if the tumour can be detected again during or after (adjuvant) therapy with curative intent. For patients, this may represent the transition to a stage of the disease that is no longer curable and is therefore directly relevant to the patient. In addition, the failure of the curative treatment approach is also relevant for patients who are not tumour-free at the start of the study and was used by IQWiG as an independent endpoint for haematooncological patients with diffuse large B-cell lymphoma (DLBCL)[20-22], but also for patients with solid tumours.[23,24]

Methodological requirements for valid surrogates

Surrogate validations are generally possible based on randomised controlled trials (RCTs) and in other special situations (as was the case with the SVR). Typically, they require a meta-analysis of multiple RCTs examining both effects on the surrogate endpoint and the patient-relevant endpoint of interest. Recognised validation methods are correlation-based methods such as the consideration of the correlation between the effects at study level and the surrogate threshold effect (STE).

In addition to the methodology used for surrogate validation, it is of fundamental importance that the underlying study pool is complete and suitable. The validation studies presented in the benefit assessments on pertuzumab (breast cancer)[16,17] and nintedanib (interstitial lung disease)[18] show that it is possible in principle to conduct a surrogate validation in accordance with IQWiG's requirements. In the pertuzumab project, the DFS and on nintedanib the annual forced vital capacity were each to be shown to be suitable surrogates for overall survival. For pertuzumab, the underlying study pool of the validation study was not suitable, as studies were excluded that would have been relevant in the therapeutic area presented in this benefit assessment. For nintedanib, the methodological implementation of the validation was flawed, which led to an underestimation of the STE. After self-calculated correction, it was shown that the effect on the surrogate was not large enough to derive an effect on overall survival.

In addition, there are special situations in which validity can also be recognised.[25] This requires that the relationship between the patient-relevant endpoint and the surrogate endpoint is clearly biologically/medically plausible and that other criteria are met. One example is the SVR in patients with chronic hepatitis C infection, where the occurrence of the surrogate endpoint led to a significantly reduced risk of hepatocellular carcinoma. In addition, the risk regarding the actual endpoint reached a minimal level, namely that of an unaffected population.[10] In special situations, cohort studies must be available as a data basis that relate to people undergoing treatment and whose

follow-up period is sufficiently long to adequately record the risk of the actual endpoint occurring.

Benefit assessment and treatment decisions require sufficiently reliable data

From IQWiG's perspective, a major issue is that regulatory authorities often accept surrogate endpoints that are not sufficiently validated. The use of such endpoints has increased over recent decades at both the Food and Drug Administration (FDA) and the European Medicines Agency (EMA).26-28 For (potentially accelerated/conditional) approval, the position is often taken that this is justifiable and that greater uncertainty is tolerated to make a new pharmaceutical quickly available.

However, the acceptance of high uncertainty at the time of approval must not lead to a situation where truly reliable data are never generated. Benefit assessment and evidence-based therapy decisions require a sufficient level of certainty. The aim of early benefit assessment is to identify new pharmaceuticals with the judgement „additional benefit proven" that are sufficiently certain to have additional value for patients and not just „maybe".

Evidence-based medical care also requires strong data for well-founded treatment decisions. This requires meaningful endpoints and, in certain cases, more or better surrogate validation. Moreover, the frequently voiced criticism that therapeutic parameters described in guidelines and disease management programmes (DMPs) are ignored by the benefit assessment also does not get to the heart of the problem: the use of a laboratory parameter recommended in a guideline or DMP for therapy management in an individual patient does not necessarily legitimise its suitability as a patient-relevant endpoint in a clinical study. The decisive factor for the benefit assessment can only be whether it is sufficiently certain that an effect in the surrogate endpoint will also be reflected in an effect in the patient-relevant endpoint.

Data situation for chronic diseases remains poor

For chronic diseases there is often still a considerable lack of suitable data for early benefit assessment. From 2021 until early August 2024, IQWiG conducted a total of 169 benefit assessments in the area of chronic diseases. No suitable data were available for 68% (115) of these projects (figure 1).

The main issue here is not primarily the endpoints, but rather the study design. Even though most of these studies were randomised controlled trials (RCTs), their data were often unsuitable for benefit assessment because these RCTs compared the intervention to a placebo or an inadequate therapy in the control arm, rather than directly comparing it with the current standard therapy. In other cases, there was no comparison at all (single-arm studies), and data from such studies are generally unsuitable for benefit assessment. It was already criticised in 2019 that there is often no suitable data available for the benefit assessment.[29] Unfortunately, no positive development can be observed regarding chronic diseases. The reasons for this stagnation and possible approaches for change (such as targeted use of positive or negative incentives) should be discussed.

Among the 54 (32%) benefit assessments conducted between 2021 and early August 2024 where suitable data were available, more than 70% were based on approval studies. In the majority of these cases, only one pivotal study was available for the benefit assessment. Non-approval studies were used for 15 (28%) of the benefit assessments (figure 1). Overall, in only about a quarter of the included studies, the primary study endpoint was fully

patient-relevant. In more than 40% of the included studies, the primary endpoint was not patient-relevant, often because it was based on a (non-sufficiently validated) surrogate endpoint, e.g. changes in HbA1c levels in type 2 diabetes or the eGFR slope in Fabry disease patients (figure 2).

With regard to eGFR, it should be noted that changes in eGFR are only patient-relevant if there is sufficiently certainty that the eGFR declines to a level that is perceptible to patients. This was the case for the first time in one of the two addenda to finerenone.30 The combined endpoint of the presented studies on renal morbidity included the individual components renal failure (defined as confirmed persistent decline in eGFR to < 15 ml/min/1.73 m2 or end-stage renal disease), eGFR decline ≥ 57% and renal death. Given the mean eGFR baseline values (approx. 43 ml/min/1.73 m2) of the patients, the ≥57% eGFR decline component was considered sufficiently patient-relevant, and the combined endpoint was used.

In 30% of the included studies, the primary endpoint could only be used with limitations (e.g. change in average monthly migraine days compared to baseline vs change in average monthly migraine days compared to baseline vs migraine days/month reduction by 50% or only individual components of a combined endpoint were considered (figure 2). The operationalisation of an endpoint can therefore be decisive for whether or not an endpoint is considered patient-relevant.

It is evident that just because an endpoint is investigated in (approval) studies does not automatically mean it is patient-relevant. Instead of „study endpoint relevant", as Thomas Kaiser already suggested in 2016,[31] the principle should rather be „relevant endpoint study endpoint". In principle, HTA decisions should not incentivise the use of study endpoints that entail a high degree of uncertainty and should instead reward the investigation of meaningful, patient-relevant endpoints.[29]

However, the question of the patient relevance of the primary endpoint or its operationalisation is rarely directly relevant to the conclusion regarding the question „Is there an additional benefit: yes or no?". Between 2021 and early August 2024, in only six projects (11%), no additional benefit was determined despite a reported advantage for the intervention in the primary (but not patient-relevant) endpoint, e.g. changed HbA1c levels in type 2 diabetes. However, in most cases, other relevant endpoints demonstrated an additional benefit (this was the case in 24 projects [44%]). A key example is the assessment of risankizumab, in which the primary endpoint (endoscopic remission) was not accepted, but positive effects were observed in the Inflammatory Bowel Disease Questionnaire (IBDQ; total score and sub-score bowel symptoms) and Short Form-36 (SF-36).[32]

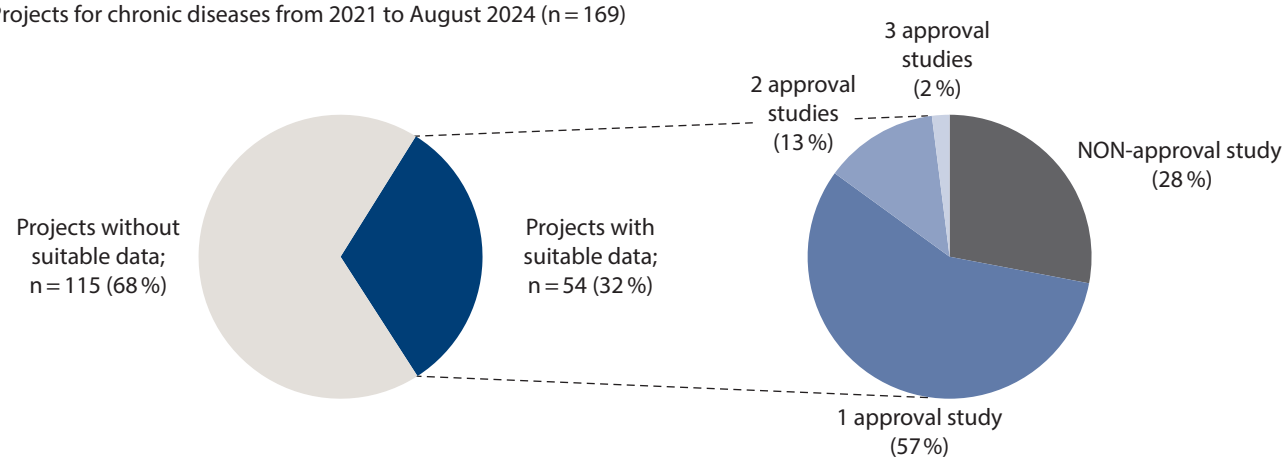
Besides positive effects, other assessments also showed negative effects that led to the overall judgement of „no additional benefit" (two projects [4%]). In seven projects (13%), the primary (non-patient-relevant) endpoint showed no effect, and the rejection had no impact on the benefit assessment conclusion (see figure 3). These studies were non-inferiority studies. However, it is problematic if in these studies lack meaningful endpoints related to mortality, morbidity, or quality of life, i.e. on benefit endpoints. This was the case, for example, for vadadustat.33 But the latest diabetes assessments (type 2) also show that the studies are not aligned with treatment goals.[34]

Improvements needed for PRO analyses

The data quality of PROs for benefit assessment is often insufficient, so that only some of the submitted PRO analyses are ultimately suitable for the benefit assessment.[35]

Chronic diseases – often no suitable data are available for the benefit assessment; if data are available, these are often approval studies

Projects for chronic diseases from 2021 to August 2024 (n = 169)



Source: IQWiG

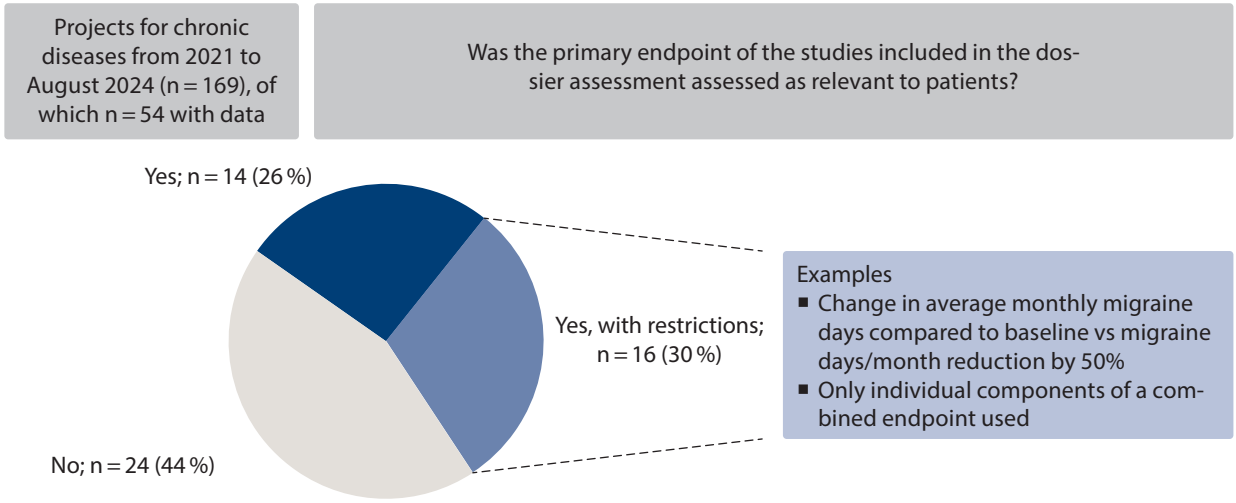
Figure 1: Of the 169 benefit assessments with reference to chronic diseases, no suitable data were available for 68% (115) of the projects in the period from the beginning of 2021 to the beginning of August 2024. The problem was not primarily the endpoints, but the study design.

Another frequent problem is that the PROs are collected too briefly. The reason for this is presumably that the studies submitted for benefit assessment are strongly tailored to regulatory approval. For example, the PROs are often only collected until disease progression. But for those affected, symptoms and quality of life remain highly relevant even after the disease has worsened. Whether the patient is better or worse off in the long term with the intervention to be evaluated than with the comparator treatment can only be determined if the corresponding PROs are recorded beyond progression.[36]

Another key factor is the completeness of PRO data collection over the study period. For example, low response rates to the corresponding questionnaires often mean that the data cannot be used for benefit assessment. A recent stakeholder discussion highlighted that patient representatives expressed a strong willingness to participate in PRO data collection, provided they are clearly informed about how their input will be used. This indicates that a longer survey of these endpoints is both feasible and does not conflict with patient interests.[37]

The analysis and operationalisation of PRO endpoints must be discussed carefully on a case-by-case basis, as numerous factors can influence the significance of results. For instance, depending on the indication and treatment goal, either a (lasting) improvement or a worsening of symptoms may be the primary focus. Should continuous analyses or responder analyses be used at a specific evaluation time point? Which time point is best suitable for a responder analysis?[38] Which response criterion is appro-

Only about a quarter of the primary study endpoints are fully relevant to patients



Source: IQWiG

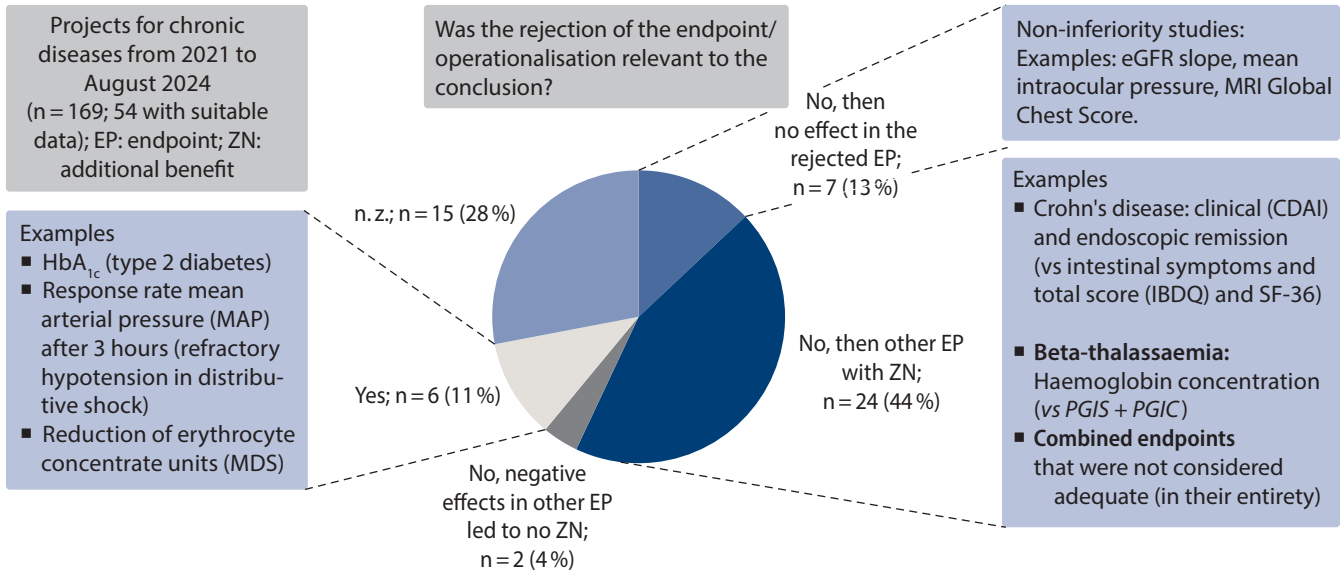
Figure 2: The operationalisation of an endpoint can also be decisive for whether or not an endpoint is considered patient relevant. The fact that an endpoint was investigated in (approval) studies does not necessarily mean that it is a patient-relevant endpoint.

priate for indicating a noticeable change? For the question of the response criterion, the patient-specific change of 15% of the range of the survey instrument proposed by IQWiG has proven to be practicable.[39] The aim of this approach developed by IQWiG was also to create clarity for manufacturers and to prevent outcome-driven reporting. Since its introduction, this response threshold has been regularly used by manufacturers in studies and dossiers.

Conclusion

- IQWiG's principles remain unchanged: patient relevance is the paradigm of benefit assessment.
- Status of surrogates: few submitted validations, mainly in oncology. Numerous publications in recent years show: regulatory authorities accept surrogates whose validity has not been adequately demonstrated – more validations (where necessary) desirable.
- Instead of „study endpoint relevant", the principle should be „relevant endpoint study endpoint"
- Improvements in PRO analyses are necessary:
  - Basic methodological quality (completeness and quality of data) & duration of PRO survey often in need of improvement (e.g. beyond progress).
  - The usefulness of analyses must be discussed on a case-by-case basis (e.g. symptom burden, [permanent] improvement/deterioration; dealing with intercurrent events; time point under consideration).
  - Overall, (even) more high-quality data on PROs is needed.

In only 11 per cent of the studies consulted was the rejection of the primary endpoint relevant to the conclusion (regarding the question additional benefit yes / no?)



CDAI: Crohn's Disease Activity Index; **eGFR:** estimated glomerular filtration rate; **IBDQ:** Inflammatory Bowel Disease Questionnaire; **MDS:** myelodysplastic syndromes; **PGIC/PGIS:** Patient Global Impression of Change or Severity; **SF-36:** Short Form-36

Source: IQWiG

Figure 3: In the vast majority of assessments, the rejection of the primary study endpoint is not relevant to the question of an additional benefit (yes or no?) due to its non-assigned patient relevance.

References

[1] Federal Ministry of Health (BMG, 2023) Pharmaceutical Products Benefit Assessment Ordinance according to § 35a Paragraph 1 SGB V for reimbursement agreements according to § 130b SGB V (Ordinance on the Benefit Assessment of Pharmaceuticals, AM-NutzenV). https://go.sn.pub/f5cp5w. Accessed 09.10.2023.

[2] Institute for Quality and Efficiency in Healthcare (IQWiG, 2023) General Methods; Version 7.0. https://go.sn.pub/mkwigq. Accessed 06.10.2023.

[3] Institute for Quality and Efficiency in Healthcare (IQWiG, 2021) Risdiplam (spinal muscular atrophy) – Benefit assessment according to § 35a SGB V; dossier assessment. https://go.sn.pub/doprfg. Accessed 11.07.2023.

[4] Institute for Quality and Efficiency in Healthcare (IQWiG, 2024) Vosoritide (achondroplasia); addendum to project A23-92 (dossier assessment). https://doi.org/10.60584/A24-08. Accessed 23 February 2024.

[5] Institute NC Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). https://go.sn.pub/n7lwyc. Accessed 07.10.2024.

[6] Institute for Quality and Efficiency in Healthcare (IQWiG, 2020) Durvalumab (non-small cell lung cancer) – Benefit assessment according to § 35a SGB V; dossier assessment. https://go.sn.pub/90yg37. Accessed 11 Juli 2023.

[7] Institute for Quality and Efficiency in Healthcare (IQWiG, 2023) Sacituzumab govitecan (breast cancer); Benefit assessment according to § 35a SGB V; dossier assessment. https://doi.org/10.60584/A23-86. Accessed 17 November 2023.

[8] Institute for Quality and Efficiency in Healthcare (IQWiG, 2012) Rilpivirine – Benefit assessment according to § 35a SGB V; dossier assessment. https://go.sn.pub/zphzud. Accessed 11 Juli 2023.

[9] Institute for Quality and Efficiency in Healthcare (IQWiG, 2015) Insulin degludec (new indication); Benefit assessment according to § 35a SGB V;

dossier assessment. https://go.sn.pub/3lai7e. Accessed 11 Juli 2023.

[10] Institute for Quality and Efficiency in Healthcare (IQWiG, 2011) Boceprevir – Benefit assessment according to § 35a SGB V; dossier assessment. https://go.sn.pub/0605wp. Accessed 11 Juli 2023.

[11] Institute for Quality and Efficiency in Healthcare (IQWiG, 2013) Dabrafenib – Benefit assessment according to § 35a SGB V; dossier assessment. https://go.sn.pub/qnv9l4. Accessed 11 Juli 2023.

[12] Institute for Quality and Efficiency in Healthcare (IQWiG, 2015) Pembrolizumab – Benefit assessment according to § 35a SGB V; dossier assessment. https://go.sn.pub/0lfwif. Accessed 11 July 2023.

[13] Institute for Quality and Efficiency in Healthcare (IQWiG, 2017) Palbociclib (breast cancer) – Benefit assessment according to § 35a SGB V; dossier assessment. https://go.sn.pub/0r0kkc.. Accessed 11 July 2023.

[14] Institute for Quality and Efficiency in Healthcare (IQWiG, 2017) Ribociclib (breast cancer) – Benefit assessment according to § 35a SGB V; dossier assessment. https://go.sn.pub/kqadrk. Accessed 11 July 2023.

[15] Institute for Quality and Efficiency in Healthcare (IQWiG, 2020) Apalutamid (prostate cancer) – Benefit assessment according to § 35a SGB V (expiry of time limit); dossier assessment. https://go.sn.pub/7dbyze. Accessed 11 July 2023.

[16] Institute for Quality and Efficiency in Healthcare (IQWiG, 2018) Pertuzumab (breast cancer) – Benefit assessment according to § 35a SGB V; dossier assessment. https://go.sn.pub/nnvllq. Accessed 11 July 2023.

[17] Institute for Quality and Efficiency in Healthcare (IQWiG, 2021) Pertuzumab/Trastuzumab (breast cancer, adjuvant) – Benefit assessment according to § 35a SGB V; dossier assessment. https://go.sn.pub/x6kivq. Accessed 11 July 2023.

[18] Institute for Quality and Efficiency in Healthcare (IQWiG, 2020) Nintedanib (other chronic progressive fibrosing interstitial lung diseases) – Benefit assessment according to § 35a SGB V; dossier assessment. https://go.sn.pub/4bj058. Accessed 11 July 2023.

[19] Institute for Quality and Efficiency in Healthcare (IQWiG, 2024) Niraparib/Abiraterone acetate (prostate cancer); addendum to project A23-107 (dossier assessment). https://go.sn.pub/g13v0o. Accessed 06 May 2024.

[20] Institute for Quality and Efficiency in Healthcare (IQWiG, 2023) Lisocabtagen maraleucel (DLBCL, HGBL, PMBCL and FL3B, second line) – Benefit assessment according to § 35a SGB V; dossier assessment. https://go.sn.pub/uz4974. Accessed 16 September 2023.

[21] Institute for Quality and Efficiency in Healthcare (IQWiG, 2023) Axicabtagene-Ciloleucel (DLBCL and HGBL, second line) – Benefit assessment according to § 35a SGB V; dossier assessment. https://go.sn.pub/w6w8p2. Accessed 10 October 2023.

[22] Institute for Quality and Efficiency in Healthcare (IQWiG, 2024) Polatuzumab vedotin (in combination with rituximab, cyclophosphamide, doxorubicin, and prednisone; previously untreated DLBCL) – Benefit assessment according to § 35a SGB V; dossier assessment. https://go.sn.pub/n4pkoh. Accessed 02 April 2024.

[23] Institute for Quality and Efficiency in Healthcare (IQWiG, 2024) Nivolumab (NSCLC, neoadjuvant); addendum to project A23-74 (dossier assessment). https://go.sn.pub/81yl50. Accessed 01 February 2024.

[24] Institute for Quality and Efficiency in Healthcare (IQWiG, 2022) Pertuzumab (breast cancer) – Benefit assessment according to § 35a SGB V; dossier assessment. https://go.sn.pub/6f2qkt. Accessed 11 July 2023.

[25] Institute for Quality and Efficiency in Healthcare (IQWiG, 2025) [A24-61] Scientific elaboration of clinical studies in the therapeutic area of wound treatment. https://go.sn.pub/t18565. Accessed 7 January 2025.

[26] Schuster Bruce C, Brhlikova P, Heath J, McGettigan P (2019) The use of validated and nonvalidated surrogate endpoints in two European Medicines Agency expedited approval pathways: A cross-sectional study of products authorised 2011-2018. PLoS Med 16(9):e1002873. doi:10.1371/journal.pmed.1002873.

[27] Gyawali B, Hey SP, Kesselheim AS (2020) Evaluating the evidence behind the surrogate measures included in the FDA's table of surrogate endpoints as supporting approval of cancer drugs. EClinicalMedicine 21:100332. doi:10.1016/j.eclinm.2020.100332.

[28] Wallach JD, Yoon S, Doernberg H (2024) Associations Between Surrogate Markers and Clinical Outcomes for Nononcologic Chronic Disease Treatments. JAMA 331(19):1646-1654. doi:10.1001/jama.2024.4175.

[29] Wieseler B, McGauran N, Kaiser T (2019) New drugs: where did we go wrong and what can we do better? BMJ 366:l4340. doi:10.1136/bmj.l4340.

[30] Institute for Quality and Efficiency in Healthcare (IQWiG, 2023) Finerenone (renal insufficiency, stage 3 and 4); addendum to project A23-15 (dossier assessment). https://go.sn.pub/qmqozn. Accessed 23 August 2023.

[31] Thomas Kaiser; Institute for Quality and Efficiency in Healthcare (IQWiG, 2016) clinically relevant endpoints and surrogates in early benefit assessment. Interdisciplinary Platform on Benefit Assessment Volume 2 January 2016

[32] Institute for Quality and Efficiency in Healthcare (IQWiG, 2023) Risankizumab (Crohn's disease); addendum to project A22-133 (dossier assessment). https://go.sn.pub/42elj9. Accessed 11 July 2023.

[33] Institute for Quality and Efficiency in Healthcare (IQWiG, 2024) Vadadustat for symptomatic anaemia due to chronic renal disease: Missed opportunity. https://go.sn.pub/a2uk9c.

[34] Institute for Quality and Efficiency in Healthcare (IQWiG, 2024) Insulin degludec (diabetes mellitus type 2); Benefit assessment according to § 35a SGB https://go.sn.pub/byfekc.

[35] Kramer L, Moos M, Thaa B (2024) Health-related quality of life (HRQoL) in German early benefit assessment: The importance of disease-specific instruments. Z Evid Fortbild Qual Gesundhwes 186:1-9. doi:10.1016/j.zefq.2024.02.003.

[36] Institute for Quality and Efficiency in Healthcare (IQWiG, 2022) Criticism of study planning: Patient-reported outcomes often collected far too briefly. https://go.sn.pub/em0o9l.

[37] Institute for Quality and Efficiency in Healthcare (IQWiG, 2023) Long-term collection of patient-reported outcomes in oncological studies: important and feasible. https://go.sn.pub/bdb35h.

[38] Pe M, Alanya A, Falk RS (2023) Setting International Standards in Analyzing Patient-Reported Outcomes and Quality of Life Endpoints in Cancer Clinical Trials-Innovative Medicines Initiative (SISAQOL-IMI): stakeholder views,

objectives, and procedures. Lancet Oncol 24(6):e270-e283. doi:10.1016/S1470-2045(23)00157-2.

[39] Institute for Quality and Efficiency in Healthcare (IQWiG, 2021) Clinical relevance of patient-reported outcomes: New threshold proves fit for practice. https://go.sn.pub/jmstw4.

Innovation in study endpoints - study endpoints for innovations

Dr Jutta Wendel-Schrief | MSD Sharp & Dohme GmbH

All over the world, clinical trials must meet the requirements of regulatory authorities as well as those related to market access and reimbursement. When considering the early benefit assessment under AMNOG, the requirements for pivotal studies often differ from those of the EMA. A crucial aspect of planning clinical trials is the selection of endpoints. The EU-HTA pursues a patient-centred HTA approach, whereas Germany takes a unique approach by defining patient-relevant endpoints. The operationalisation of morbidity endpoints requires a clear definition of clinical relevance and consideration of the heterogeneity of the patient population. A consensus between medical societies, clinical researchers, and patient representatives is essential to formulate appropriate endpoints and collect relevant quality of life data. MSD is actively involved the development of both clinically and HTA-relevant endpoints. EU-HTA provides an opportunity for harmonisation and further development of AMNOG towards greater patient centricity.

Pharmaceutical companies as a source of innovation in Germany and worldwide

For decades, pharmaceutical companies have played a significant role in the development of innovative pharmaceuticals that advance healthcare globally. According to a study by the European Federation of Pharmaceutical Industries and Associations (EFPIA), the pharmaceutical industry invested almost 50 billion Euros in research and development in Europe in 2023, with approximately half (48.4%) of these funds directed towards clinical trials.

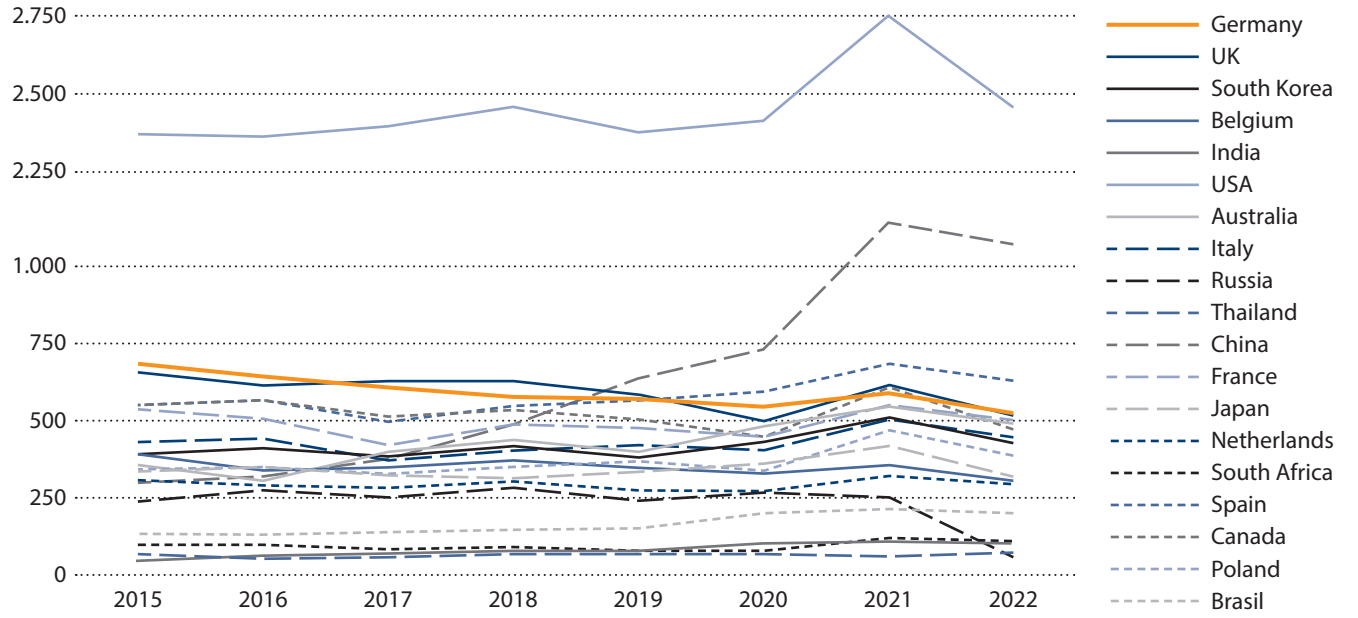
In Germany alone, the pharmaceutical industry invested around 7 billion Euros in research and development in 2020, ranking sixth worldwide in conducting industry-initiated clinical trials, with 542 clinical trials. According to an analysis by Charles Rivers Associates, the private sector accounts for almost two-thirds of the investments in research and development.

A major challenge for pharmaceutical companies: balancing different requirements

‚The design of clinical trials repeatedly presents significant challenges for sponsors. Clinical trials must not only meet the requirements of medical science but also those of regulatory and reimbursement authorities. For a globally operating company like MSD Sharp & Dohme (MSD), which markets medicines in over 140 countries, this means that these requirements must be considered to ensure regulatory approval and market access. Despite the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) and the European HTA regulation/EUnetHTA, significant differences in standards remain.

At MSD, approximately one third of employees worldwide work in research and development, currently

Number of clinical trials conducted by pharmaceutical companies in a country comparison

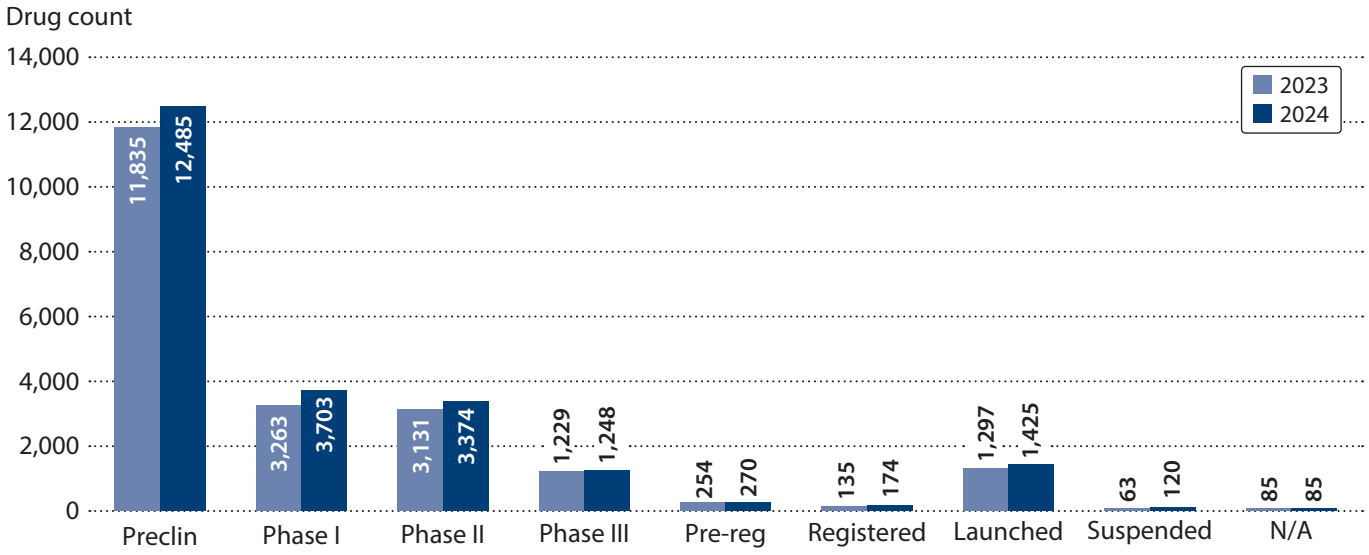


Source: vfa based on clinicaltrials.org
see https://www.vfa.de/de/arzneimittel-forschung/so-funktioniert-pharmaforschung/amf-standortfaktoren.html

Figure 1: In 2020, Germany ranked sixth worldwide in the number of industry-sponsored clinical trials, with 542 studies.

Dr Jutta Wendel-Schrief *has been Head of the Market Access division at MSD Sharp & Dohme GmbH since January 2015. She studied biology and completed her doctorate in biochemistry at the Johann Wolfgang Goethe University in Frankfurt am Main. Following several years of research at the Max Planck Institute for Biophysics on the subject of transport proteins, she worked in various pharmaceutical companies since 1990, covering different roles and indications, primarily in marketing and sales. From 2010 to 2015, she headed the Clinical and Speciality Products business unit at MSD Sharp & Dohme GmbH.*

Pipeline of pharmaceuticals in development, comparison 2024 to 2023



N/A = not applicable and is applied to companion diagnostics prelaunch

Source: Pipeline by development phase, Pharmaprojects January 2024

Figure 2: On average, only one to two out of every 10,000 pharmaceuticals in development reach market maturity.

engaged in planning and conducting over 2,000 clinical trials across 47 countries and four regions, including 130 clinical trials in Germany, with recruitment for 20 additional studies underway. A large number of internal and external experts, including clinicians, operations specialists, research partners, and patients, are involved in drawing up the study protocols. The goal is to create feasible study protocols that comply with global requirements while also addressing German-specific regulations.

Despite these efforts, on average, only one to two out of every 10,000 pharmaceuticals in development reach market maturity.

Operationalisation of morbidity endpoints in the benefit assessment

Under the AMNOG, the efficacy, tolerability, usefulness, and harmfulness of therapies are measured based on patient-relevant endpoints, which are assessed using the categories of mortality, morbidity, side effects and health-related quality of life. The key question is which endpoints are considered patient-relevant and how they can be operationalised. There has been considerable interpretative flexibility, particularly regarding morbidity endpoints.

1. Clinical relevance: definitions must be precise and meaningful to appropriately measure the intended clinical effects.

2. Heterogeneity of the patient population: endpoints

must adequately capture the diversity of patients and their varying disease courses (consideration of subgroups).

3. Consensus on endpoints: the definition of morbidity endpoints requires consensus between different stakeholders, including medical societies, researchers, regulatory authorities, HTA bodies, and patient representatives.

4. Data collection and availability: the operationalisation of morbidity endpoints must be feasible in everyday medical practice to ensure that the necessary data for assessing these endpoints can be collected, especially regarding quality of life.

5. Patient perspective: it is crucial to appropriately consider the patients' perspectives, needs, and preferences when defining and evaluating patient-relevant endpoints.

Achieving consensus on suitable endpoints is not always possible and can be time-consuming and resource intensive.

Under the AMNOG, an additional challenge arises from the fact that the G-BA's requirements for the operationalisation and measurement of patient-relevant endpoints do not always fully reflect the clinical research question. As a result, pharmaceuticals that fill an important treatment gap for patients may not demonstrate an additional benefit under AMNOG and could therefore become unavailable in Germany.

Example of event-free survival (EFS)

A key example is the endpoint event-free survival (EFS), which captures disease progression and relapse, among other aspects. In the curative treatment situation, the IQWiG and G-BA view EFS as an indicator of failed potential cure, which is generally considered patient-relevant. For the acceptance of this endpoint, the G-BA requires proof of tumour control or tumour-free status. However, for inoperable patients, tumour assessment (tumour control) during treatment is often challenging due to radiotherapy-induced tissue changes in the tumour.

The questions posed in the benefit assessment therefore do not reflect those of the clinicians, which in this example means that the G-BA's requirements cannot be implemented for the pharmaceutical companies. Closer coordination between the G-BA, medical societies, patient representatives and clinical research is essential to address these issues. The goal must be to develop clinically relevant and practicable morbidity endpoints. The need for this is high – especially for inoperable patients with curative treatment intentions.

The phenomenon of the discrepancy between the G-BA criteria and the reality of everyday clinical practice also exists for operable patients. The G-BA requires proof of tumour-free status (R0 resection) for EFS acceptance, meaning that all patients who are not completely resected are counted as events in the EFS analyses. According to IQWiG and G-BA, the curative approach is considered unsuccessful for these patients – even though they may still have a chance of being cured.

When developing the study protocol, the challenging question for the company is whether the operationalisation of the outcome should be close to everyday care and thus adequately reflect the varying disease progression of patients or whether the focus should be on the success of the benefit assessment when recording the outcome. This raises the question of whether the G-BA's requirements for such endpoints should be reconsidered to better align with clinical practice and prioritise patient needs.

Addressing these challenges requires careful consideration and collaboration between all stakeholders, including regulatory authorities, the pharmaceutical industry, medical societies, and patient representatives, to ensure that

AMNOG benefit assessments are appropriate and comprehensive. This is the only way to ensure that the data of patients who have invested time, commitment and hope in studies can be used in a meaningful way. Unassessable datasets are a waste of valuable resources for both pharmaceutical companies and patients.

Patient-relevant endpoints and innovations of tomorrow

Tomorrow's innovations present the system with new challenges. For example, MSD is collaborating with Moderna to develop individualised neoantigen therapies (INTs) – also known as „therapeutic cancer vaccines". The aim of INT is to activate the body's own immune system in patients who already have cancer in such a way that it can help fight the cancer in order to minimise the risk of recurrence. They are designed to help the immune system recognise and destroy individual neoantigens on the patient's cancer cells. The current research results indicate that the INT immune cells could be activated in such a way that they can recognise and attack the altered neoantigens on the cancer cells. Due to the immune system's memory, INTs could lead to long-term, specific recognition of the cancer cells and reduce the risk of recurrence. In the V940-001 trial, INT is being investigated regarding the primary endpoint of recurrence-free survival (RFS). Secondary endpoints include distant metastasis-free survival (DMFS), overall survival (OS) and safety of therapy.

Operationalisation of symptom-related progression-free survival (PFS) (PFS+) using artificial intelligence

An important endpoint in oncology in some indication areas is the PFS endpoint, which is not recognised per se as patient-relevant in AMNOG. A further development of the endpoint (PFS+), which links PFS with disease symptoms or quality of life, could be a way to achieve acceptance of the endpoint in the early benefit assessment, as it would demonstrate the direct influence of progression on the patient's well-being. This would make PFS+ patient-relevant in the sense of the AMNOG. Artificial intelligence could support this process by identifying patterns in existing datasets that establish this connection.

Will EU-HTA make AMNOG endpoints more patient-centred?

Medical progress is advancing rapidly, and in some cases, randomised controlled trials (RCTs) are therefore not feasible for practical and ethically reasons. This dynamic must be reflected in the benefit assessment of pharmaceuticals to prevent regulatory frameworks from lagging behind scientific advancements. For special therapeutic situations, such as gene and cell therapies, as well as INTs, for which studies of the highest evidence level are impossible or inappropriate, the AMNOG therefore needs to be further developed. With the introduction of European HTA assessment in January 2025, there is an opportunity to align methodological assessment principles more closely with the European framework, ensuring that medical innovation continues to have a place in Germany. More flexibility is needed regarding the consideration of patient-centred endpoints and specific study designs (e.g. indirect comparisons) in line with European guidelines so that medical progress also has a chance in Germany in the future.

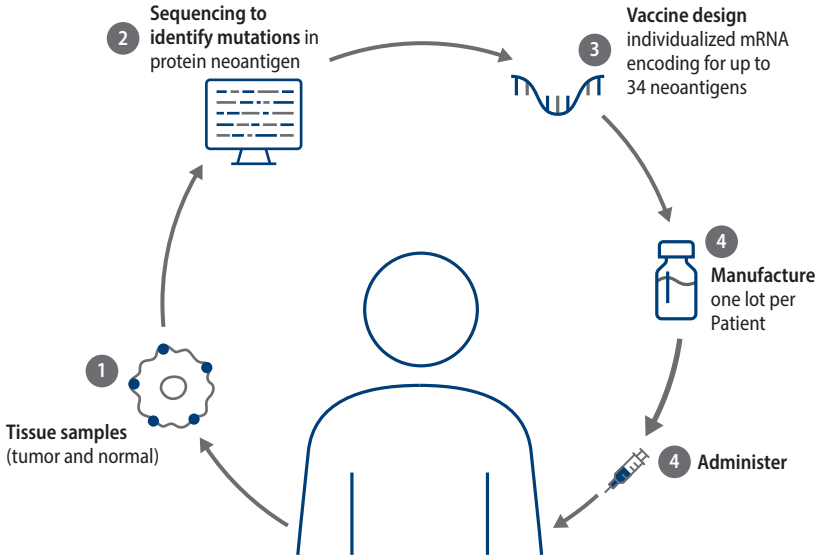
References

[1] The Pharmaceutical Industry in Figures, EFPIA, edition 2024, see the-pharmaceutical-industry-in-figures-2024.pdf, page 5.

[2] vfa based on clinicaltrials.org, see https://go.sn.pub/7yka34.

[3] Fowler A, Grieve K, Maos A, Wilsdon T. Quantifying public and private investment in European biopharmaceutical research and development. Health Aff Sch. 2024 May 31, see Quantifying public and private investment in European biopharmaceutical research and development - PubMed page 2

Individualised neoantigen therapy – Study with focus on the primary endpoint of recurrence-free survival (RFS)



Advantages of ‚RNA-based Cancer Therapies

Speed and efficiency
Needle-to-needle generation of personalized treatments can be produced in just a few weeks.

Integrated Manufacturing
Opportunities to scale and optimize within Moderna's existing Massachusetts manufacturing facility

Customizable
Identify and encode multiple patient-specific neoantigens to design an individualized treatment

Source: Moderna

Figure 3: The aim of INT is to activate the body's own immune system in patients who already have cancer in such a way that it can help fight the cancer to minimise the risk of recurrence.

[4] https://go.sn.pub/xdyrkx.

[5] Clinical trials by MSD – msd.de

[6] Citeline Clinical, Pharma R&D Annual Review 2024, see https://go.sn.pub/1pfji1. Figure 2, page 9

Patient-centred outcomes in the EU-HTA versus AM benefit regulation and IQWiG methods paper

	EU: patient-centered outcomes	AM-NutzenV: (& rules of procedure)	IQWiG. Methods paper version 7.0
Mortality	✓	✓	✓
Morbidity	✓	✓	✓
QoL	✓	✓	✓
Side effects/Safety perception	✓	✓	✓
(feelings)	✓	✗	✓
Beliefs Preferences	✓	✗	✗
Präferenzen	✓	✗	✗
Needs	✓	✗	✗
Functions and activities	✓	✗	✓

Source: vfa

Figure 4: The introduction of the European HTA assessment in January 2025 presents an opportunity to better align methodological assessment principles with the European framework.

Patient-relevant endpoints: Questions from the perspective of medical services

Dr Andreas Rhode | The Medical Service of the health funds in Westphalia-Lippe

s the emphasis on mortality in benefit assessment justified?

Patient-relevant endpoints analyse mortality, morbidity, and quality of life. In many diseases that were previously rapidly fatal, medical progress has led to an improvement in patient survival. Moreover, the use of additional pharmaceuticals administered in later therapy lines falsifies statements about overall survival. In these situations, life extension is neither the primary therapeutic goal nor a meaningful study endpoint. Event- and progression-free survival are now considered clinically meaningful endpoints as surrogate parameters, with analyses demonstrating a correlation with overall survival (Michael Untch et al., Disease-free survival (DFS) as a surrogate for overall survival (OS) in patients (pts) with HR+/HER2– early breast cancer (EBC): A correlation analysis. JCO 41, 535-535(2023). DOI:10.1200/JCO.2023.41.16_suppl.535). However, there are also numerous examples in which an advantage in progression-free survival (PFS) was ultimately not significantly reflected in overall survival.

Does the mortality criterion in the benefit assessment have to be revised?

Surrogate parameters as a meaningful measurement tool in benefit assessment

For many diseases, both mortality and morbidity cannot be effectively controlled within the time frame of a study.

Can surrogate parameters (e. g. PFS, DFS, but also laboratory chemical parameters such as HbA1c etc.) be used in this context?

What requirements must these surrogate parameters fulfil so that conclusions can actually be drawn about patient-relevant parameters?

Patient-relevant endpoints as secondary endpoints

In the development of pharmaceuticals, phase III studies are designed for approval and not for benefit assessment. Consequently, the parameters that might later be assessed in the benefit assessment were only set as secondary endpoints (e.g. primary endpoint PFS, secondary endpoint mortality, quality of life, etc.).

How biased are the results regarding patient-relevant endpoints if these only represented the secondary endpoint?

Patient-relevant endpoints in indirect comparisons

Since pharmaceutical companies often select studies for approval that do not represent a direct comparison with an appropriate comparator therapy (ACT), as this is not required by the regulatory authorities and the ACT can also differ in the various countries, the benefit assessment procedure do not include direct comparisons with the ACT. Pharmaceutical companies often choose indirect comparisons. Here, two studies are analysed with each other by means of a bridge comparator. However, the parameters for the defined endpoints may differ between the studies.

How can indirect comparisons be managed in relation to patient-relevant endpoints?

What are the minimum requirements for comparability?

Rigid patient-relevant endpoints versus highly individualised treatment response

Bamberger (2020) reports a highly individualised treatment response for epilepsy or depression, for example. There are no therapy predictors for neurological and mental illnesses that could be used to define patient collectives. As a result, an additional benefit cannot be demonstrated (Bamberger M. Welchen patientenrelevanten Nutzen haben neue Arzneimittel in der Neurologie und

Psychiatrie? (What patient-relevant benefits do new drugs in neurology and psychiatry offer?) Psychopharmakotherapie 27:289–294, 2020.).

Are the patient-relevant endpoints set too rigidly in the benefit assessments? Does this potentially undermine the clinical added value of pharmaceuticals?

Dr Andreas Rhode *is a specialist in psychiatry and psychotherapy and a certified health economist. After several years as a senior physician at St. Marien Hospital Hamm (Chair of Psychotherapy and Psychosomatics at the University of Witten/Herdecke), he took over the management of the specialist clinic Release in 2011. Since 2021, he has led the Social Medicine Expert Group 6 for Supply of Pharmaceuticals within the Medical Services of Health Funds community and serves as an advisor to the GKV-SV in the Subcommittee for Pharmaceuticals of the G-BA. He is also a member of the Off-Label Expert Group and the Long COVID Off-Label Expert Group at the BfArM.*

Collection of patient-reported outcomes: insights from psychiatry

Professor Peter Falkai, Clinic for Psychiatry and Psychotherapy, University Hospital of Munich | Professor Dr Jörg Ruof, European Access Academy, Basel

Modern psychiatry focussed on patient well-being is essentially based on differentiated patient-reported outcome measures (PROMs, Patient Reported Outcome Measures). However, the dominance of a single endpoint such as overall survival in oncology is not foreseeable for psychiatric disorders due to their complex progression. Instead, various dimensions of patient-reported outcomes must be equally considered, such as generic instruments; disease-specific instruments; and the collection of PREMs (Patient-Reported Experience Measures), which capture the patient's experience with healthcare services. Breakthrough innovations are currently less foreseeable for psychiatric disorders. Research efforts are therefore focused on gradually improving the care of psychiatric patients based on the collection of PROMs and PREMs, as well as relevant surrogate parameters such as „paid employment" and „stable partnership".

Significance of patient-reported indicators in psychiatry

The European regulation on Health Technology Assessment (EU HTAR) aims to strengthen the European Health Union.[1] In the medium to long term, discussions on endpoints for clinical studies and patient care will be further intensified and standardised. Patient-reported outcomes will play a central role in this process. The EU HTA Coordination Group defines Patient-Reported Outcomes (PROs) in its methodological guidance as „any report of the status of the patient's health condition that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else".[2]

For psychiatric disorders, these assessment instruments play an important role in both clinical practice and research. Since 2017, a working group within the OECD has been working on the development of international standards for patient-reported indicators for mental illnesses.[3] The harmonisation of PRO assessment instruments is essential for sharing experiences across national borders and defining international endpoints for conditions such as depression, bipolar disorders, or schizophrenia.

The systematic integration of PRO instruments into psychiatric care enables, for example, the early detection of warning signs in patients at risk of suicide, or the collection of other important „between-visit" data in outpatient or day-care settings.[4]

However, the effective use of these instruments requires expertise and experience.[4] Patient cooperation and motivation are crucial, as is the „buy-in" of staff in hospitals or outpatient care facilities.

Differentiation between PROM and PREM

The terms PRO or PROM (Patient Reported Outcome Mea-

sures) refer to various concepts that need to be considered in a differentiated manner. A systematic review analysed the outcome of lifestyle interventions, such as a healthy diet and physical activity, in patients with severe psychiatric disorders. The meta-analysis included 21 studies with a total of 5,907 patients. There was no effect of the interventions mentioned on quality of life. However, disease-specific instruments measuring the severity of depression or anxiety symptoms showed a positive effect.[5] Conceptually, generic assessment tools with a broad and comprehensive perspective, should be differentiated from disease-specific assessment instruments focussed on the specific disease. Moreover, the Patient-Reported Experience Measures (PREM) must be differentiated from the generic or disease-related PROMs. These are also reported by the patients themselves but are less focussed on the patients' own health status and more on their experiences with healthcare.[6, 7]

Definition of PROMs and PREMs

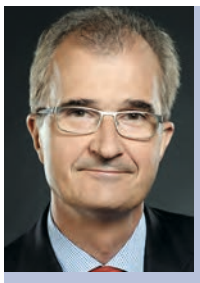
For decades, clinician-rated outcome measures have been the central source of data informing clinical practice and policy

- Patient Reported Outcome Measures (PROMs) more directly assess the lived experiences of service users, capturing their perspectives on their health status and essential subjective constructs such as goal attainment, quality of life and social inclusion
- Patient Reported Experience Measures (PREMs) assess their experiences of using health services, including communication, responsiveness and recovery orientation

Source: Distinction between PROM and PREM according to Roe (2022)[7]

Figure 1: A distinction must be made between PROMs and PREMs, which focus primarily on the patient's experience.

A study conducted at Université Paris Cité and published in 2023 collected PROMs and PREMs from 248 patients



Professor Peter Falkai *has been Chair of the Department of Psychiatry and Psychotherapy at the LMU Munich Hospital since 2012. From 2011 to 2012, he was President of the German Society for Psychiatry, Psychotherapy and Neurology and is a member of the Leopoldina National Academy of Sciences.*

Professor Jörg Ruof *is a member of the Advisory Board of the Benefit Assessment Platform. He is a specialist in general medicine, holds a Master of Business Administration and a Master of Public Health. He is associate professor at the Institute for Epidemiology, Social Medicine and Health Systems Research at Hannover Medical School. In 2017, he founded r-connect GmbH and in 2021, the „European Access Academy".*

admitted to a university psychiatric hospital. All three dimensions: i) generic PROMs; ii) disease-specific PROMs; and iii) PREMs were collected. There was an improvement in PROMs during hospitalisation. The PREMs closely matched the parallel ratings of the physicians but showed only weak correlations with established clinical PRO measurement instruments. Accordingly, the authors recommend measuring all three dimensions as indicators of therapeutic success and the quality of patient care.[8]

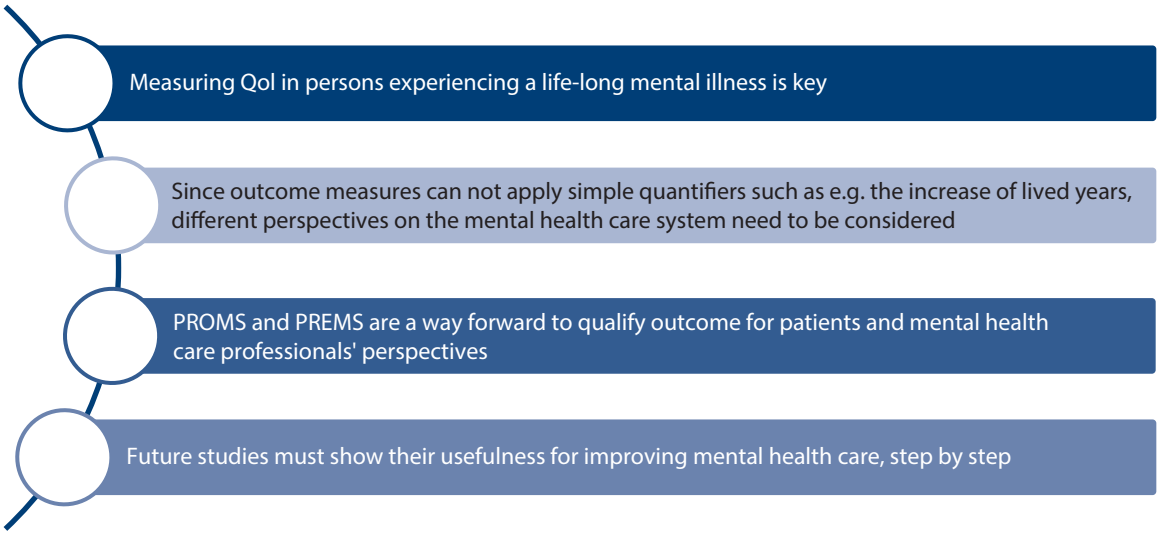
Trends in patient-centred healthcare

Current trends towards „user-centred healthcare models" and „shared decision-making" highlight the relevance of patient-centred and partnership-based healthcare. Ethical considerations, the consideration of patients' rights and, above all, the focus on better therapeutic outcomes through active and equal patient involvement make these trends irreversible. The integration of different perspectives (patients, relatives, caregivers) is just as important as the consideration of the various dimensions of patient-reported outcomes.

A survey of patients and the Medical Service of Health Insurance (MDK)[9] revealed:

- Mortality is generally not a suitable outcome for psychiatric disorders (except in suicide studies).
- Relapse-free and (fully) functional long-term survival (recovery concept), on the other hand, are a meaningful endpoint.

Overview of the status of outcome measurement in psychiatry



Source: Own presentation

Figure 2: The dominance of a single endpoint is not foreseeable in psychiatry. Instead, various dimensions of patient-reported outcomes must be equally considered.

- Surrogate parameters for this include paid employment and stable partnerships (secondary endpoints). Primary endpoints are also the measurement of psychopathology and functioning.

PROs are another important way of primarily testing the efficacy of pharmaceuticals or other therapeutic interventions

In summary, it should be emphasised that modern psychiatry, which is oriented towards patient well-being, is essentially based on differentiated patient-centred assessment instruments. The dominance of a single endpoint, such as overall survival in oncology, is neither foreseeable nor likely to be appropriate given the complexity of psychiatric disorder.

Instead, the various dimensions of patient-reported outcomes must be equally considered, such as generic broad-based instruments; disease-specific instruments; and the collection of PREMs, i.e. the patient's experience with healthcare services. Breakthrough innovations are currently less foreseeable. Research efforts are therefore focused on gradually improving the care of psychiatric patients based on the collection of PROMs and PREMs, as well as relevant surrogate parameters such as „paid employment" and „stable partnership".

References

[1] 1 Regulation (EU) 2021/2282 of the European Parliament and of the Council of 15 December 2021 on health technology assessment and amending Directive 2011/24/EU. https://go.sn.pub/kp5l9f Accessed 25 January 2025

[2] HTA Coordination Group: Guidance on outcomes for joint clinical assessments. Adopted on 10 June 2024 by the HTA CG pursuant to Article 3(7), point (d), of Regulation (EU) 2021/2282 on Health Technology Assessment https://go.sn.pub/qahh23. Accessed 25 January 2025

[3] de Bienassis K, Kristensen S, Hewlett E, et al. Patient-reported indicators in mental health care: towards international standards among members of the OECD. Int J Qual Health Care 2021; 34 (Suppl 1): ii7-ii12.

[4] Fowler JC, Madan A, Frueh BC et al. Lessons learned while integrating patient-reported outcomes in a Psychiatric Hospital. Psychotherapy 2019; 56: 91-99

[5] Pape LM, Adriaanse MC, Kol J et al. Patient-reported outcomes of lifestyle interventions in patients with severe mental illness: a systematic review and meta-analysis. BMC Psychiatry 2022; 22: 261 https://doi.org/10.1186/s12888-022-03854-x. Accessed 25 January 2025

[6] de Bienassis K, Kristensen S, Hewlett E, et al. Measuring patient voice matters: setting the scene for patient-reported indicators. Int J Qual Health Care 2021; 34 (Suppl 1): ii3-ii6.

[7] Roe D, Slade M, Jones N. The utility of patient-reported outcome measures in mental health. World Psychiatry. 2022 Feb;21(1):56-57

[8] Scanferla E, de Bienassis K, Pachoud B, Gorwood P. How subjective well-being, patient-reported clinical improvement (PROMs) and experience of care (PREMs) relate in an acute psychiatric care setting? European Psychiatry 2023; 66: e26 1-9 https://doi.org/10.1192/j.eurpsy.2023.12. Accessed 25 January 2025

[9] Personal communication with Dr Andreas Rhode, Medical Service Westphalia-Lippe

Endpoints – challenges from the vfa's perspective

Dr Andrej Rasch | Association of Research-Based Pharmaceutical Companies (vfa)

In the context of the AMNOG benefit assessment, there are numerous practical and methodological challenges in dealing with endpoints such as overall survival, PROs, or surrogate endpoints. In addition to the question of fundamental relevance, there is also a considerable need for discussion regarding the weighting and overall consideration of individual outcomes and effects. A stronger focus is required on accepted and established methods that comply with international standards of evidence-based medicine, along with greater transparency in the classification and evaluation of endpoints, as well as consideration of the specificities of therapeutic situations.

„Which targets are relevant for patients?" This is the key question of the 2024 autumn meeting of the Interdisciplinary Platform for Benefit Assessment. The question of endpoints is one of the key issues in the AMNOG procedure. But how important are the endpoints for determining an additional benefit in the AMNOG procedure? From a hierarchical point of view, it must be objectively stated that the primary factor is whether the study design is compatible with the AMNOG procedure (i.e. routinely randomised controlled) of sufficient duration and the correct implementation of the specification for the appropriate comparator therapy. However, if these key conditions are met, the actual focus shifts to the endpoints and the study effects demonstrated for them.

From a legal perspective, Section 2 para. 3 AM-NutzenV specifies that the „benefit of a pharmaceutical [...] is the patient-relevant therapeutic effect, particularly concerning the improvement of health status, reduction of disease duration, prolongation of survival, reduction of side effects, or improvement of quality of life." Therefore, additional benefit must be determined based on the impact on „patient-relevant endpoints". The status of an endpoint in approval-relevant studies (primary or secondary) also has no influence on the relevance in the benefit assessment.

After approximately 13 years of assessment practice, approximately 57% of new pharmaceuticals were able to demonstrate their additional benefit. However, this proportion varies significantly depending on the therapeutic area. While an additional benefit was demonstrated for oncological pharmaceuticals in approximately 73% of cases, the proportion of proven additional benefit for diseases of the nervous system or psychiatric disorders

was 46% and only 29%, respectively. Obviously, the combination of a compatible study design and endpoints led to better results for oncological pharmaceuticals in the AMNOG procedure than in other indications.

Overall survival

It should be noted that, the endpoint „overall survival" is of paramount importance, particularly in oncology. It is undoubtedly patient-relevant and therefore also decisive for the demonstration and classification of additional benefit. The empirical evidence for oncological diseases shows a clear dependence between the magnitude of the benefit in overall survival (in combination with benefits in other endpoint categories) and the extent of additional benefit in the decisions of the G-BA.

However, there are also several practical and methodological challenges for the „overall survival" endpoint. For example, it is not possible to achieve an appropriate study power within a reasonable time frame in all treatment situations. This is further complicated in the benefit assessment when a study population is divided into sub-questions by the G-BA or only part of the pivotal study population is considered as usable. This is done without any recognisable consideration of the power problem.

In some situations, a permitted change of therapy is also ethically unavoidable, which makes it difficult to interpret the survival data. Although there are methodological solutions for dealing with the authorised change of therapy, none of these approaches have yet been accepted in the AMNOG assessment. Equally challenging are the requirements for capturing follow-up therapies in interpreting overall survival results.

If a particular therapeutic situation does not (yet) allow any statements to be made about overall survival within a reasonable time frame, the question also arises as to how a patient-relevant treatment success can be measured. In the decision-making practice of the G-BA, endpoints such as EFS (event-free survival), DFS (disease-free survival), or RFS (relapse-free survival) are now generally accepted but only to illustrate the failure of a curative treatment approach.

However, the PFS (progression-free survival) endpoint is still not relevant to the assessment, although there have been differing opinions on PFS within the G-BA from the outset with regard to patient relevance. Endpoints such as TTST (time to subsequent therapy), CR (complete response, except for the previous special case of basal cell carcinoma assessment) or MRD (minimal residual disease) are also not taken into account. These endpoints, which are important or even primary for authorisation decisions, are consistently classified in the AMNOG procedure as non-patient

Dr PH Andrej Rasch *has been working as Senior Manager Benefit Assessment/HTA Coordination at the German Association of Research-Based Pharmaceutical Companies (vfa) since 2013. Prior to that, he was Head of the Pharmaceuticals Research Department at the Scientific Institute of the AOK (WIdO), a methodologist at the Federal Joint Committee (G-BA) and a research associate at the Faculty of Health Sciences at Bielefeld University.*

relevant and thus non-assessable, regardless of the specifics of individual therapeutic situations.

Surrogate endpoints

In addition to the question of direct patient relevance, for some endpoints the question arises as to whether they can be used as substitutes (surrogates) for other endpoints in certain therapeutic situations. In AMNOG benefit assessments, the IQWiG guidelines published in 2011 with the Rapid Report A10-05 „Significance of surrogate endpoints in oncology" apply to surrogate validation (IQWiG 2011).

Ideally, a surrogate validation involves a meta-analysis of several RCTs with high result certainty, a strong correlation measure at both study and individual levels, or alternatively, the application of the surrogate threshold effect (STE) concept with specific threshold values. Although the G-BA requirements do not specify explicit thresholds, they refer to the same methodology as proposed by IQWiG. So much for the claim. The reality is also sobering, as these requirements have not been met for 13 years now. Whether this is due to the conservative nature of the requirements themselves, or a limited number of validation attempts remains unanswered.

Nevertheless, it should be noted that surrogate endpoints are indispensable in some cases, as certain key issues cannot otherwise be investigated to enable access to innovative treatments. In the past, some surrogate endpoints were therefore also accepted in the AMNOG benefit assessment in exceptional cases, even without formal validation according to the above-mentioned methods, and used to derive the added benefit.

For example, the endpoint of sustained virologic response in chronic hepatitis C, virologic response in HIV infection or HbA1c in type 1 diabetes mellitus were categorised as sufficiently valid surrogate endpoints by both IQWiG and the G-BA. While considering these surrogate endpoints was always comprehensible, the respective decision was still based on a less than transparent individual assessment.

Patient-reported outcomes

Patient-reported outcomes (PROs) are increasingly gaining significance. In many therapeutic areas, capturing morbidity and health-related quality of life through these endpoints has become a standard in clinical studies. A review of regulatory approvals in the EU for oncological pharmaceuticals between 2017 and 2020 showed that PROs were included in approximately 78% of pivotal studies (Teixeira et al. 2022).

The increasing importance of PROs is also evident in the AMNOG benefit assessment. For instance, in the case of non-small cell lung carcinoma, 95% of the studies considered by the G-BA provided usable data on at least one PRO instrument (Brand et al. 2022). The picture for the subgroup of patient-reported outcomes, health-related quality of life, is also encouraging. The proportion of procedures with data on quality of life has increased in recent years and has been over 70% since 2014. This proportion was particularly high for benefit assessments on oncological pharmaceuticals (Kramer et al. 2024).

The question of fundamental relevance does not end with the type of an endpoint but can also extend to its operationalisation. For example, a PRO outcome that appears to be clearly relevant may not be considered in an assessment. This can be illustrated using the example of benefit assessments in the therapeutic area of moderate-to-severe plaque psoriasis. Here, the PASI 90 endpoint, which represents a 90% improvement in disease symptoms and almost symptom-free skin, has not been considered by IQWiG for years, as it cannot be formally ruled out that psoriasis symptoms are still present and affect

patients. For this reason, IQWiG exclusively relies on evaluations of PASI 100 (complete remission). From the outset, this assessment was at odds with the guidelines and healthcare practice, where PASI 75 and PASI 90 responses also serve as treatment targets, as the absence of cutaneous symptoms cannot be achieved in all patients (Nast et al. 2021). Consequently, the G-BA also considers the corresponding results for PASI 75 and PASI 90.

There are also several challenges regarding the collection of PROs. For example, there are no validated and established instruments available for some special therapeutic situations, such as rare diseases. The use of existing questionnaires from other therapeutic areas is generally viewed critically. In the interpretation of study results, potential power issues remain unconsidered. It is also challenging to record and maintain high response rates, especially in terminal phases of life and after progression of a life-threatening disease (Böhme et al 2022).

Until recently, there were differing views on the duration of PRO recording. On the one hand, there were differences between IQWiG, which advocates documentation for as long as possible until the end of the study, and the clinical experts, who believe that recording after progression is important, but to a reasonable extent and not without restrictions until the end of life.

Dealing with available evidence

With regard to the basic acceptance of the data, reference should be made to the existing provision in Section 5 (5) AM-Nutzen, which states: „If valid data on patient-relevant endpoints are not yet available at the time of assessment, the evaluation shall be based on the available evidence, taking into account the quality of the studies, indicating the likelihood of proof of additional benefit, and a deadline may be set by which valid data on patient-relevant endpoints must be submitted."

On the one hand, the regulation is aimed at the possibility of a time limit, which is already common practice. On the other hand, it mandates that evaluations should be based on available evidence. In practice, however, it has been shown that available data are generally not used if they are not categorised according to patient-relevant outcomes. This raises the question of whether an assessment should be carried out taking into account the available evidence, especially in special therapeutic situations.

Weighting of endpoints and effects

In addition to the fundamental question of the relevance of an endpoint, the question of how relevant an outcome or effect is also arises in the context of a benefit assessment. IQWiG distinguishes between three hierarchical categories of outcome in its own methodology: 1. overall mortality; 2. serious (or severe) symptoms and side effects and health-related quality of life, and 3. non-serious (or non-severe) symptoms and side effects.

However, the classification of a target measure as severe or non-severe is not always sufficiently transparent or straightforward. Using the example of an application area such as moderate to severe plaque psoriasis outlined above, it can be observed that a blanket classification of the endpoint PASI 100 (complete remission) under „non-serious/non-severe symptoms" can certainly raise questions. In many cases, the hierarchical classification of an outcome (e.g., from an EORTC QLQ-C30 questionnaire for oncological diseases) is based solely on its formal classification into the categories of morbidity or quality of life. This can lead to a systematic bias in endpoint classification in the morbidity category, as these may then be categorised as „non-severe" and therefore have a higher hurdle in the assessment.

Moreover, the methodology for determining the extent of additional benefit also raises several critical questions. This special approach to assessing effect size has been controversial from the outset, particularly due to the predefined threshold values for upper confidence interval limits, normative determinations, or the assumption of two studies across all therapeutic situations.

Although the G-BA has not relied on the IQWiG methodology for determining the extent of benefit since 2011 (a fact explicitly stated in the rationale of all resolutions), it is nevertheless assumed that this methodology continues to have a lasting influence on benefit assessments. The established threshold values for continuous endpoints, in combination with the conservative approach of a shifted hypothesis threshold, do not align with internationally recognised criteria or standards of evidence-based medicine and thus represent an additional challenge (IQWiG 2022).

Another issue is the evaluation of the fundamental relevance of PRO (patient-reported outcome) effects. The requirement for established and validated MID (minimal important difference) thresholds has been replaced in responder analyses by a rigid 15% formula. Accordingly, if responder analyses are pre-specified in a study and the response criterion corresponds to at least 15% of the scale range of the assessment instrument used, these analyses are taken into account in the evaluation.

However, this „one-size-fits-all" approach is controversial for multiple reasons, particularly as it also represents a deviation from the international scientific approach to improving assessment standards through meaningful quality criteria. It further fails to sufficiently consider the known differences in patient perspectives regarding meaningful outcomes (Böhme et al. 2022, Schlichting et al. 2022).

Moreover, the IQWiG methodology leads to a situation where even if clinical relevance is ensured through the predefined responder criterion, a statistically significant effect for some PROs does not necessarily result in recognition as a meaningful effect. This is because, in addition to the aforementioned response criterion of 15%, another relevance criterion applies: the threshold value for the upper confidence interval (for non-severe symptoms). This leads to a duplication of the relevance criteria and an over-conservative categorisation of the PRO effects.

Overall assessment

Ultimately, the overall assessment of endpoints and therapeutic effects is central to decisions regarding additional benefit. The G-BA conducts this assessment on behalf of patients and their preferences. However, the weighting of these factors lacks a formal and sufficiently transparent procedure. Studies on measuring patient preferences have not yet been considered within the AMNOG procedure.

Some classifications raise questions here, e.g. in the case of categorisation of the therapeutic benefits as minor additional benefit. According to the AM-NutzenV, this is the case if a „previously unachieved moderate and not merely minor improvement in the therapy-relevant benefit [...] is achieved, in particular a reduction in non-serious symptoms of the disease or a relevant avoidance of side effects".

In the G-BA's assessment practice, however, this also includes assessments with a prolongation of overall survival, avoidance of relapses in oncological diseases, more frequent complete remissions of severe plaque psoriasis in children and adolescents or multiple benefits in patients with moderate to severe active Crohn's disease. With the introduction of the so-called „guidelines" in the

GKV-FinStG, the evaluation of effects has become even more critical, as even a small change in the classification of the prolongation of additional benefit – particularly when incorporating methodological uncertainties – can determine whether it falls within the scope of these guardrails in subsequent negotiations.

European perspective

The European HTA process is also a prospective and yet imminent challenge. This will start in January 2025 with the evaluation of advanced therapy medicinal products (ATMPs) and oncological medicinal products. From 2028 onwards assessments of orphan drugs and from 2030 for other pharmaceuticals will follow. A major uncertainty is the number of national PICO questions, which depend on requested and available endpoints, as well as possible operationalisations and evaluations of the endpoints.

This also raises the challenge of how the national „Delta-Dossier" for the AMNOG procedure will be structured. Additionally, it remains to be seen whether the desired harmonisation of methodological requirements will be achieved in the future and what interactions will result from the different handling of endpoints in the European HTA and the AMNOG benefit assessment. For example, with regard to the anchored approach of „patient-centred endpoints" (including e.g. preferences or needs), which is pursued in the European HTA.

Conclusion

In conclusion, key challenges in dealing with endpoints in the context of benefit assessments can be summarised as follows. There is a need for:

- a stronger focus on accepted and established methods that meet the international standards of evidence-based medicine,
- greater transparency in the categorisation and weighing up of endpoints,
- consideration of the special features of therapeutic situations in the benefit assessment.

References

[1] Böhme et al. Es braucht neue Ansätze für Lebensqualitätsdaten (New Approaches for Quality of Life Data Needed), in Monitor Versorgungsforschung (01/22), p. 43-47.

[2] Brand et al. Value in Health, Volume 25, Issue 12S (December 2022), https://go.sn.pub/y32ial.

[3] IQWiG. Aussagekraft von Surrogatendpunkten in der Onkologie. (Significance of surrogate endpoints in oncology). Rapid Report. IQWiG-Reports – Year: 2011 No. 80

[4] IQWiG. Dokumentation und Würdigung der Anhörung zum Entwurf der Allgemeinen Methoden 6.1 (Documentation and consideration of the hearing on the draft General Methods 6.1). Version 1.0 dated 22 March 2022

[5] Kramer et al. Health-related quality of life (HRQoL) in German early benefit assessment: The importance of disease-specific instruments, in: ZEFQ (2024), https://doi.org/10.1016/j.zefq.2024.02.003

[6] Nast et al. German S3 Guideline for the Treatment of Psoriasis Vulgaris, Adapted from EuroGuiDerm – Part 1: Therapy Goals and Recommendations. J Dtsch Dermatol Ges 2021 Jun;19(6):934-951.

[7] Schlichting et al. Is IQWiG's 15% Threshold Universally Applicable in Assessing the Clinical Relevance of Patient-Reported Outcomes Changes? An ISPOR Special Interest Group Report. Value Health 2022 Sep;25(9):1463-1468.

[8] Teixeira et al. (2022) A review of patient-reported outcomes used for regulatory approval of oncology medicinal products in the European Union between 2017 and 2020. Front. Med. 9:968272.

Patient-relevant endpoints in the context of the European HTA process

Dr Johanna Seeger | Federal Joint Committee (G-BA)

The first joint European assessments (Joint Clinical Assessments, JCA) and consultations are scheduled to commence in January 2025. These require clearly defined methodological requirements and a common understanding of key concepts. Insights gained from practice exercises to determine the scope of the assessment within the so called „scoping" procedure has been incorporated into the development of various guidance documents. Regarding endpoints, the „Guidance on Outcomes for JCA" plays a significant role in harmonising requirements for endpoint within the framework of the JCA. However, the authority to interpret the relevance of outcomes for specific endpoints remains with the member states.

Introduction

With the entry into force of Regulation (EU) 2021/2282 (EU HTA Regulation) on 12 January 2025, the first joint European assessments (Joint Clinical Assessments, JCA) and consultations (Joint Scientific Consultations, JSC) for pharmaceuticals are imminent. The scope of the EU HTA Regulation will initially be limited to pharmaceuticals for the treatment of oncological diseases and advanced therapy medicinal products (ATMPs), but will be gradually extended to include all new pharmaceuticals from 13 January 2030.[1]

A prerequisite for implementing the EU HTA Regulation is the development of a common understanding among EU member states and the definition of common methodological requirements as a basis for the implementation of JCAs and JSCs. For this purpose, guidance documents on various procedural and methodological aspects are being developed within the subgroups of the HTA Coordination Group.

The backbone of JSCs and JCAs is the PICO scheme, which defines the research questions of European HTA assessments through the parameters „population, intervention, comparator, outcomes". A suitable operationalisation of endpoints is crucial for the generation of meaningful study data for the JCA process and thus also represents a fundamental aspect of consultation in the JSCs.

Endpoints in the scoping process

The determination of the assessment scope for a JCA is conducted during the so-called „scoping" procedure. In this process, the national research questions from member states are gathered based on PICO schemes. These are then consolidated to address the needs of individual member states with as few PICOs as possible.

This process was tested as part of the „EUnetHTA 21" project using three example pharmaceuticals.[2] The results of these „PICO exercises" formed the basis for the further development of the „Guidance on the Scoping Process", intended to provide direction for member states as well as assessors, and co-assessors in the definition of national PICOs and subsequent consolidation.

The PICO exercises revealed significant heterogeneity in the formulation of endpoints among member states (table 1).

First of all, it is noticeable that the usual categorisation of the endpoints in Germany into mortality, morbidity, health-related quality of life, and side effects was not conducted. Moreover, the individual endpoints varied greatly in terms of their level of specificity. In some cases, specific measurement instruments were mentioned, in others they were formulated in very general terms: „any other patient-centred outcome measured by patient-reported outcome measures". Furthermore, in certain cases there was a mixture of data collection methods and endpoints („radiological tumour assessment, including overall response rate and duration of response"). For safety endpoints, the possibility of defining an indication-independent, standardised set became apparent.

The experience gained from the PICO exercises regarding the formulation of endpoints was incorporated into the „Guidance on Outcomes for Joint Clinical Assessments (JCA)".[3]

Dr Johanna Seeger *has worked as an advisor in the Pharmaceuticals Department of the German Federal Joint Committee (G-BA) since 2021. Since 2022, she has represented the G-BA in the subgroups of the HTA Coordination Group for joint clinical assessments, for joint scientific consultations and for the development of methodological and procedural guidelines. In 2009, she graduated from the Protestant University of Applied Sciences Berlin with a Bachelor of Nursing degree. From 2012 to 2016, she studied pharmacy at the FU Berlin. She obtained her licence to practise as a pharmacist in 2017 and completed her doctorate in 2021.*

Guidance on outcomes for JCA: Specification of endpoint requirements

This guidance document, developed by the subgroup for the development of methodological and procedural guidelines, defines key concepts and requirements for endpoints in the JCA process. It also provides guidelines for how member states should formulate endpoints as part of the scoping process. Standardised wording is intended to support the consolidation of endpoints. For instance, the guidance recommends not requiring effect measures. However, if member states require specific effect measures or measurement tools to answer their national questions, it is recommended that a preference be stated with the wording „[Outcome of interest] measured preferably as [insert measure]".

The guidance document also provides instructions on handling surrogate and combined endpoints in the scoping process, as well as the related requirements for the European dossier and the JCA report. Regarding the validity, reliability, and interpretability of measurement

Results of the EUnetHTA 21 "PICO exercises" on the pharmaceuticals Pluvicto and Pombiliti on the PICO parameter "Outcomes"

Lutetium (^{177}Lu) Vipivotidtetraxetan (Pluvicto)	Cipaglucosidase alfa (Pombiliti)
Pluvicto in combination with androgen deprivation therapy (ADT) with or without androgen receptor (AR) pathway inhibition is indicated for the treatment of adult patients with progressive prostate-specific membrane antigen (PSMA)-positive castration-resistant prostata cancer (mCRPC) who have been treated with AR pathway inhibition and taxane based chemotherapy.	Pombiliti (cipaglucosidase alfa) is a long-term enzyme replacement therapy used in combination with the enzyme stabiliser miglustat for the treatment of adults with late-onset Pompe disease (acid α-glucosidase [GAA] deficiency).
■ Overall survival ■ Radiological tumor assessment, including overall response rate and duration of response ■ Progression free survival (radiological, clinical or PSA) by investigator and blinded independent committee review ■ Symptomatic skeletal event, including time to first skeletal event ■ Prostate specific antigen levels ■ Pain measured by a patient-reported outcome measure such as a numeric rating scale or a visual analogue scale ■ Fatigue ■ Health-related quality of life, measured preferably by generic and disease specific questionnaires, ie EORTC QLQ C30 plus, if possible, EORTC PR25 or FACT-P, FACT-G ■ Health status measured preferably by EQ-5D-5L ■ Any other patient centred outcome measured by patient-reported outcomes measures ■ Adverse events (total) ■ Serious adverse events ■ Severe adverse events (Grade ≥ 3) ■ Discontinuation and interruption due to adverse events ■ Adverse events of special interest (AESI) ■ Suspected unexpected serious adverse reaction (SUSAR)	■ Overall survival ■ Ventilator-free survival ■ Changes in mobility (incl. measurement by 6MWT and documented use of wheelchair) ■ Changes in respiratory function (incl. measurement by FVC in sitting and upright positions) ■ Changes in muscle strength (by validated scales) ■ Changes in motor function (by validated scales, e.g. quick motor function test) ■ Respiratory symptomatology associated with Pompe disease ■ Gastrointestinal symptomatology associated with Pompe disease ■ Quality of life (as assessed using disease-specific (preferably) and/or generic questionnaires) ■ Health status (measured preferably by the EQ-5D) ■ Patient-reported outcomes to include R-PAct scale, and any other patient-centered outcome assessed by means of a patient-reported outcome measure ■ Adverse events (AEs) (incl. hypersensitivity, infusion reactions, immunogenicity) ■ Serious AEs (SAEs) ■ Severe AEs ■ Discontinuation and interruption of treatment due to AEs ■ Mortality due to AEs

6MWT: 6-minute walking test; AE: Adverse event; EQ-5D: EuroQoL five-dimension scale questionnaire; FVC: Forced vital capacity; R-PAct: Rasch-built Pompe-specific Activity.

Source: EUnetHTA 21: D5.4 JCA without HTD submission (PICO exercises); https://www.eunethta.eu/d5-4/

Table 1: The process for minimising the number of PICOs was tested in the PICO exercises. This revealed great heterogeneity in the formulation of endpoints among member states.

instruments, the guidance specifies which information must be included in the dossier and outlined in the JCA report to enable member states to assess the suitability of a measurement tool. Finally, a standardised set of endpoints in the adverse event category is defined, for which results must always be presented in the dossier and JCA report, irrespective of the submitted and consolidated PICO schemes.

It is important to note that JCAs do not involve evaluative decisions, such as assessing the patient relevance of specific endpoints or the validity of surrogate endpoints or measurement tools. The evaluation of the (patient) relevance of endpoints, as well as conclusions drawn from the results presented in the JCA report, and associated evaluative decisions, remain the responsibility of the member states.

Conclusion

While some open questions remain just weeks before the start of the first JCA procedures, the intensive discussion of the different perspectives of the member states through the preparatory work of the HTA Coordination Group subgroups has achieved considerable progress toward a common understanding. The „Guidance on Outcomes for JCA" makes a significant contribution to harmonising the requirements for endpoints. This should minimise the duplication of identical or almost identical endpoints in the scoping process and the resulting duplication of the presentation of results in the European dossier and the JCA report.

The decision as to which specific endpoints are included in the assessment scope and which endpoints are considered relevant for answering the national questions lies with the member states. The process of determining the assessment scope for a JCA may result in outcomes for endpoints required by other member states being presented in the JCA report for a PICO used to answer the German research question, but which are not considered relevant by the German Federal Joint Committee (G-BA). The presentation of results on an outcome in the JCA report therefore does not result in an obligation to use these results for the national assessment.

Regarding the AMNOG procedure, the principle of equal treatment of products falling within the scope of the EU HTA Regulation and those assessed exclusively at national level is particularly relevant. This is especially important since there will initially be parallel procedures, depending on the indication, with and without upstream JCA, where no different assessment standards should be applied.

References

[1] Regulation (EU) 2021/2282 of the European Parliament and of the Council of 15 December 2021 on health technology assessment and amending Directive 2011/24/EU. https://go.sn.pub/bomyql.

[2] EUnetHTA 21: D5.4 JCA without HTD submission (PICO exercises). https://go.sn.pub/4jvktg.

[3] European Commission: Guidance on Outcomes for Joint Clinical Assessments. https://go.sn.pub/1pvyzz.

A review of the internet sources was conducted on 15 November 2024.

Validation of surrogate outcomes

Prof. Dr Tomasz Burzykowski / Professor of Biostatistics and Statistical Bioinformatics at the Data Science Institute of Hasselt University

A surrogate outcome (or endpoint) is intended to replace a clinical outcome for the evaluation of new treatments when it can be measured more cheaply, more conveniently, more frequently, or earlier than that clinical outcome. A surrogate outcome is expected to predict clinical benefit, harm, or lack of these. A quantitative assessment of the strength of evidence for surrogacy requires the demonstration of the prognostic value of the surrogate for the clinical outcome, and evidence that treatment effects on the surrogate reliably predict treatment effects on the clinical outcome. The process of generating the evidence is called the validation of the surrogate. Towards this aim, the so called meta-analytic approach is most often used that entails analysis of data from multiple randomized clinical trials.

The „Guidance on outcomes for joint clinical assessments" (HTA CG, 2024), adopted by the Member State (MS) Coordination Group (CG) on Health Technology Assessment (HTA), mentions that „Surrogate outcomes may or may not reflect a direct patient-centred benefit and their clinical relevance and fit to the joint clinical assessment (JCA) need to be considered by MSs." What are surrogate outcomes? As stated in the guidance: „A surrogate outcome is an outcome that is intended to replace an outcome of interest that cannot be observed in a specific clinical study. (…) A surrogate outcome is expected to only predict the treatment effect of an outcome that is not observed in a clinical study." When discussing the use of surrogate outcomes, the guidance states that they „(…) can be requested in addition to patient-centred outcomes where relevant. However, only surrogate outcomes for which validity has previously been clearly established should be requested where possible." Thus, the concept of a „validity of a surrogate outcome" is brought forward.

The validity is linked to „the strength of the association between the surrogate outcome and the outcome of interest and the association of treatment effects on the surrogate and the outcome of interest", as the guideline states that the health technology developer should „demonstrate the strength" in case a surrogate is to be applied. The requirement is compatible with the framework defined by the so called meta-analytic approach to validation of surrogate endpoints (Buyse et al., 2000).

In fact, the approach is explicitly mentioned in the guideline: „There are several other useful approaches for the validation of surrogate outcomes. In general, these methods are based on a meta-analytic approach." It is worth noting that the guideline also explicitly mentions a validity measure that is specific to the meta analytic

approach, i.e., the surrogate threshold effect: „The concept of the surrogate threshold effect is helpful for decision-making because it represents the minimum effect regarding the surrogate outcome that is required to conclude that there is also high certainty of an effect on the patient-centred outcome".

In what follows, we explain in more details the concepts and methods mentioned in the HTA CG guideline.

In the meta analytic approach (Buyse et al., 2000), data from historical clinical trials are used to evaluate the strength of the association between treatment effects on the outcome of interest and on the surrogate outcome. This is termed the trial-level association. Its strength is captured by using the correlation coefficient R or the coefficient of determination R2 obtained from a linear regression model fitted to the estimated treatment effects. The stronger the association, the more precise prediction of the treatment effect on the outcome of interest based on the effect on the surrogate. The focus on the strength of the association results from the fact that, as it has been mentioned earlier, the prediction is the main goal of the intended use of a surrogate.

Additionally, in the meta analytic approach, data from historical clinical trials are used to evaluate the strength of the association between the outcome of interest and the surrogate outcome. This is termed the individual-level association. It is usually captured by using an association measure (e.g., correlation coefficient, odds ratio) based on a bivariate model that is fitted to the individual patient data while taking into account possible treatment effects. The stronger the association, the more precise prediction of the outcome of interest based on the value of the surrogate observed for a patient.

A surrogate valid at the individual level may be useful for, e.g., tailoring patient's treatment needs (Buyse et al., 2022). However, for such a surrogate, it does not immediately follow that treatment-induced changes on the surrogate will be strongly associated with corresponding changes on the clinical outcome, i.e., that the surrogate will be valid at the trial level. For the latter case, a direct evaluation of the strength of the trial level association is needed (Buyse, 2016; Buyse et al., 2022).

To illustrate the concepts and issues related to the validation of surrogate outcomes, we consider three examples of such a validation by using the meta-analytic approach. In particular, we discuss two meta-analyses of randomized gastric-cancer clinical trials and a meta-analysis of randomized colorectal-cancer trials.

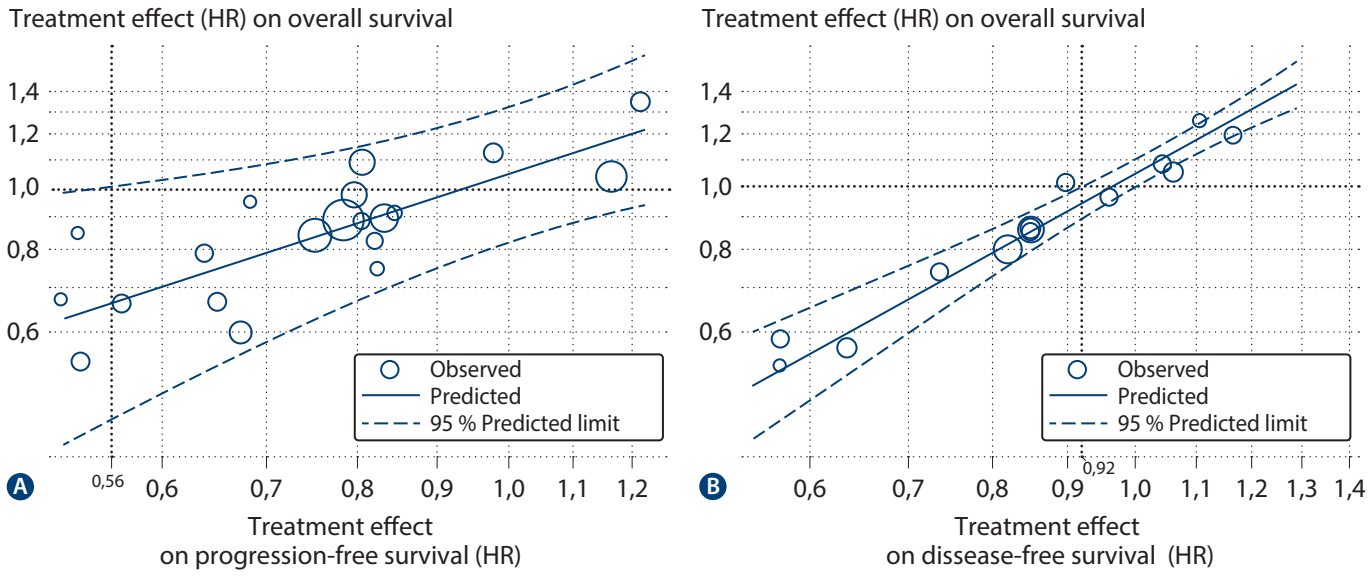
The first meta-analysis of gastric cancer trials included patients with advanced cancer (Paoletti et al., 2013). It was

Dr Tomasz Burzykowski *is Professor of Biostatistics and Statistical Bioinformatics at the Data Science Institute of Hasselt University (Belgium). He is also Vice-President of Research at the International Drug Development Institute (IDDI), a contract research organization providing clinical trials services. Tomasz held Visiting Professor posts at the Karolinska Institute (Sweden), Warsaw University of Technology (Poland), and the Medical University of Bialystok (Poland). In 2023-2024, he served as the President of the International Society for Clinical Biostatistics.*

used for the purposes of evaluating progression free survival (PFS) as a surrogate for overall survival (OS). Data were available on 4,069 patients from 20 eligible randomized trials with documented OS and PFS. The trials were investigating addition of experimental chemotherapeutic agents to pre-existing control or standard regimens. Spearman's rank correlation coefficient, quantifying the individual-level association between OS and PFS, was equal to 0.853 (95% CI [0.852, 0.854]). As mentioned in the HTA CG guideline (HTA CG, 2024), „there is no universally accepted threshold for the establishment of sufficient correlations between the surrogate and the patient-centred outcome at trial level (i.e., correlation of the effects) and patient level (i.e., correlation of the outcomes). However, a correlation of at least 0.85 is described as ‚high' and can be used as a criterion for validation of surrogate outcomes." If we adopt the threshold of 0.85, PFS is a valid individual level surrogate for OS.

The trial-level association between the treatment effects on PFS and on OS is presented in Figure 1A. The linear

Association at study level in advanced (Fig. 1a) and resected (Fig. 1b) gastric carcinoma



Source: 1a: Paoletti et al. (2013); 1b: Oba et al. (2013

Figure 1. Panel A: trial-level association for the advanced gastric cancer example. Panel B: trial-level association for the resected gastric cancer example. Circles (with size proportional to the sample size of the trial) represent the pairs of the estimated treatment effects on the surrogate and overall survival. The solid straight line is the linear regression providing the predicted treatment effect on overall survival. The dashed vertical line indicates the surrogate threshold effect. Note that both axes are on a log scale.

regression model was ln(HROS) = 0.042 + 0.779 x ln(HRPFS) where ln(HROS) and ln(HRPFS) are logarithms of the hazard ratio (HR) for OS and PFS, respectively. After adjusting for the estimation error, R2 was estimated to be equal to 0.61 (95% CI [0.04, 1.00]). The large confidence interval reflects the uncertainty around this estimate, due in part to the small sample sizes of some of the trials included in the meta-analysis. The correlation coefficient corresponding to R2 is R=0.781. By using the threshold of 0.85, PFS is not a valid trial level surrogate for OS.

The results were externally validated by using 12 trials not included in the meta-analysis, for which treatment effects were extracted from reports published in the literature after the meta-analysis was completed. Figure 2A shows the same regression line as in Figure 1A. The observed treatment effects on survival (HROS) are shown for these 12 trials, as well as the treatment effects on OS predicted from the treatment effects on the surrogate (HRPFS) in these trials, along with their 95% prediction intervals. As can be seen in Figure 2A, the prediction intervals were wide and included the value of 1 (no treatment effect on OS) for all trials, which means that the observed effects on PFS would not have allowed to predict a significant effect on OS in any of the 12 trials. Yet, three of the 12 trials showed a statistically significant effect of treatment on survival (Paoletti et al., 2013). All in all, PFS does not appear to be a valid trial level surrogate for OS in advanced gastric cancer.

The second meta-analysis of gastric cancer trials included patients with resected cancer (Oba et al., 2013). It was used for the purposes of evaluating disease free survival (DFS) as a surrogate for OS. Data were available on 3,371 patients from 14 trials with documented OS and DFS. The trials were comparing adjuvant chemotherapy with surgery alone. Based on the individual patient data, Spearman's rank correlation coefficient, quantifying the individual-level association between OS and DFS, was estimated to be equal to 0.974 (95% CI [0.971, 0.976]). By using the threshold of 0.85, DFS is a valid individual level surrogate.

At the trial level, there was also a tight association between the treatment effects on DFS and on OS (see Figure 1B). The linear regression model was ln(HROS) = 0.047 + 1.239 x ln(HRDFS). After adjusting for the estimation error, R2 was estimated to be almost equal to 1 (95% CI [0.999, 1.000]), with the corresponding value of the correlation coefficient R also close to 1. (It is worth noting that because the estimated value of R2 is very close to the upper limit of 1, the obtained numerical results need to be treated with caution.) Thus, one could conclude that that DFS is a valid trial level surrogate for OS.

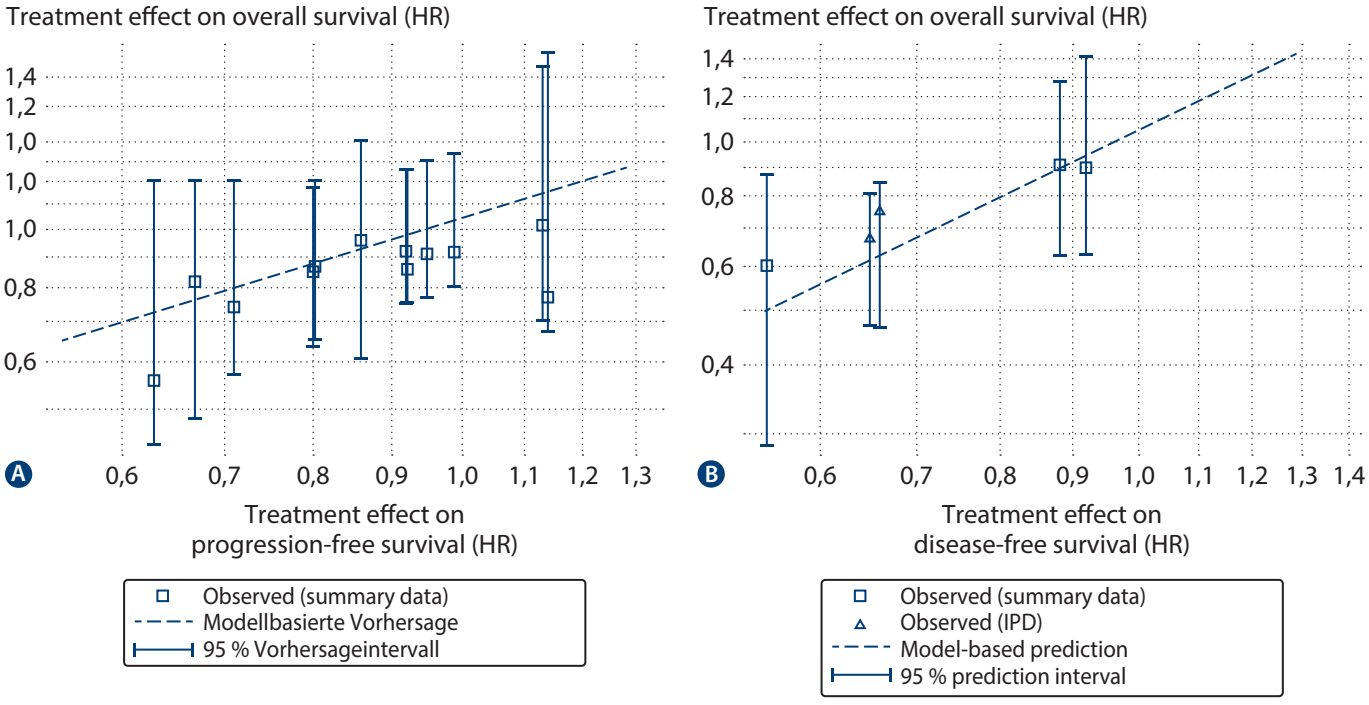
The results were externally validated by using five trials not included in the meta-analysis, three for which the treatment effects were extracted from reports published in the literature, and two for which individual patient data became available after the meta-analysis was completed. Figure 2B shows the same regression line as in Figure 1B. The observed treatment effects on survival (HROS) are shown for these five trials, as well as the treatment effects on survival predicted from the treatment effects on the surrogate (HRDFS) in these trials, along with their 95% prediction intervals. There is very good agreement between the observed and the predicted treatment effects, and in the three trials for which the prediction limits of HROS excluded one, the observed effects on survival actually reached statistical significance (Paoletti et al., 2013). All in all, DFS is a valid individual and trial level surrogate for OS in resectable gastric cancer.

It is worth noting that Figure 1 illustrates the link, mentioned earlier, between the strength of the trial level association and the precision of the prediction of the treat-

ment effect on the clinical outcome. In Figure 1A, the prediction limits for the effect on OS are wide; recall that, in this case, the value of R2=0.61 (with R=0.781) was considerably smaller than 1. On the other hand, the prediction limits in Figure 1B are narrow; in this case, R2 (and R) was very close to 1. Thus, the higher R2 or R, the more precise prediction. This is the reason why, in the meta analytic approach, the trial level validity of a surrogate is assessed by quantifying the strength of the trial level association. A similar argument applies to the individual level association.

The prediction limits, presented in Figure 1, illustrate also the concept of the surrogate threshold effect (STE). The STE is the minimum treatment effect on the surrogate that allows predicting a (non zero) effect on the clinical outcome. In practical terms, for the considered examples,

Validation of the surrogates progression-free and disease-free survival in advanced (Fig. 2a) and resected (Fig. 2b) gastric carcinoma



Source: 2a: Paoletti et al. (2013); 2b: Oba et al. (2013)

Figure 2. Panel A: validation data for the advanced gastric cancer example. Panel B: validation data for the resected gastric cancer example. The dashed straight line is the linear regression providing the predicted treatment effect on overall survival. The boxes are the estimated treatment effects on overall survival for the validation trials. The vertical intervals are the 95% prediction intervals for the treatment effect on overall survival. Note that both axes are on a log scale.

the STE is the value of HR for surrogate such that the upper prediction limit derived from the regression line becomes equal to 1. In Figure 1, the value is indicated by the dashed vertical line. For the advanced gastric cancer example, the STE was equal to 0.56. It means that a value of HR for PFS smaller than 0.56 would predict, with 95% confidence, a value of HR for OS smaller than 1. The STE of 0.56 implies a large, 44% reduction of the hazard. If we consider the range of values of HR for PFS observed in trials presented in Figure 1A, HR=0.56 represents a rather extreme treatment effect. This is a consequence of the wide prediction limits seen in Figure 1A. Thus, PFS would not seem to be a useful surrogate, because one would require a large effect on it in order to be able to claim an effect on OS. The conclusion agrees with the one derived earlier from the value of R2.

On the other hand, for the resected gastric cancer example, the STE was equal to 0.92 (see Figure 1B). It is a small treatment effect, implying only 8% reduction of the hazard. This small STE is a consequence of the narrow prediction limits seen in Figure 1B. Thus, DFS is a useful surrogate, because one would require a small effect on it in order to be able to claim an effect on OS. This conclusion coincides with the one obtained based on the value of R2.

The two gastric cancer examples illustrate how a quantitative evaluation can inform the use of surrogate outcomes; even though these two situations appear to be relatively similar, we come to different conclusions about the use of PFS and DFS as surrogates for OS. Indeed, one can argue that DFS can be used as a reasonable surrogate for OS in the adjuvant setting, while PFS cannot be used reliably as a surrogate in advanced disease.

The meta-analysis of trials for patients with advanced colorectal cancer was used for the purposes of evaluating PFS as a surrogate for OS (Buyse et al., 2007). Data were available on 4,352 patients from 13 trials. The set of trials included 10 historical trials that compared fluouracil (FU) combined with leucovorin with either FU alone (1,744 patients) or with raltitrexed (1,345 patients), and three validation trials comparing FU + leucovorin with or without irinotecan or oxaliplatin (1,263 patients). Spearman's rank correlation coefficient, quantifying the individual-level association between OS and PFS, was equal to 0.82 (95% CI [0.82, 0.83]). Strictly speaking, the value is smaller than 0.85, which could be taken as an indication PFS may not be a valid individual level surrogate.

At the trial level, there was a tight association between the treatment effects on PFS and on OS (see Figure 3). The linear regression model was ln(HROS) = 0.003 + 0.81 x ln(HRPFS). The value of R2 was estimated to be equal to 0.98 (95% CI [0.88, 1.08]), with the corresponding R=0.99. By using the threshold of 0.85, PFS is a valid trial level surrogate for OS.

The results were externally validated by using three (more recent) trials not included in the meta-analysis. The predicted effects agreed extremely well with the observed effects in trials testing irinotecan, but less well in the trial testing oxaliplatin, in which the predicted effect overestimated the observed effect (see Figure 3). The difference could be a result of the effect of second-line treatments and crossovers, which were more available as compared to the historical trials. In this respect it is worth noting that PFS is observed under first-line therapy, while OS is observed under subsequent-line therapies that may confound the effect of first-line therapy on the clinical outcome.

This may explain why PFS would be an acceptable surrogate for OS in advanced colorectal cancer in an era of marginally effective FU-based therapies (Buyse et al., 2007), while it is a much less convincing surrogate today, because patients who are in progressive disease may receive many lines of active therapy that may further impact their survi-

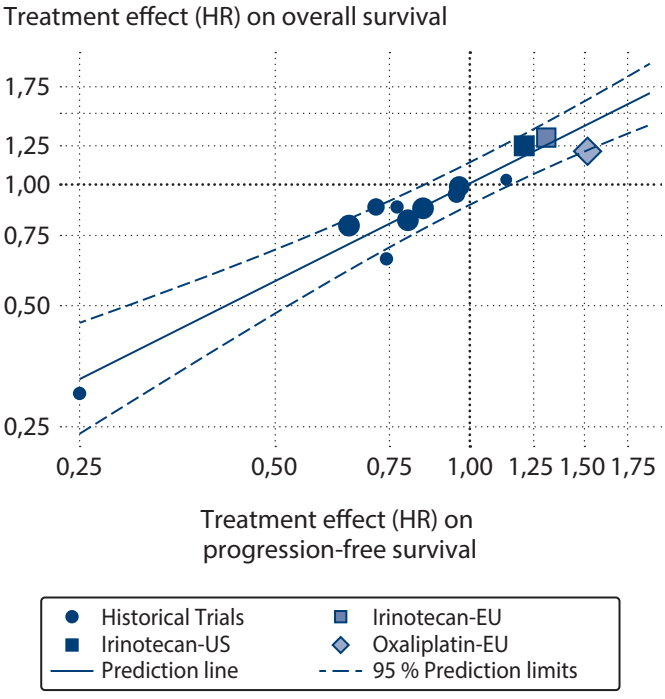
val (Shi et al., 2014).

Thus, the colorectal cancer example illustrates that the validity of a surrogate may change over time. Moreover, when compared to the advanced-gastric cancer case, it also shows that the validity of the same surrogate (PFS) may be different for different diseases. It follows that one cannot simply infer the validity of a surrogate from a validation exercise conducted in another disease (Buyse, 2016). A related, and perhaps the most challenging, question is whether a surrogate that has been evaluated for a given class of drugs is still likely to be valid for different classes of drugs.

In most cases, a combination of biological reasoning and statistical evidence will be required for a surrogate to be used outside of the conditions in which it was initially evaluated. Statistical evidence favoring use of the surrogate will be most helpful for new drugs that are similar to the drugs used in the evaluation datasets. For new drugs having a substantially different mode of action, whether the surrogate can be used in confidence is an open question that may warrant another prospective evaluation.

From a practical point of view, the meta analytic approach to validation of surrogate outcomes requires availability of individual patient data from multiple randomized clinical trials, in which information about the surrogate and clinical outcomes is available. Despite the existing data-sharing initiatives, obtaining such data still poses a challenge. Moreover, even if access to such data can be obtained, it is not guaranteed that the data will contain information about, for instance, patient reported outcomes (like quality-of-life measurements) that might be the outcome of interest for HTA. These practical issues, combined with the conceptual issues mentioned earlier, make validation of candidate surrogate outcomes still a challenging task.

Association of progression-free survival in advanced colorectal cancer



Source: Buyse et al. (2007)

Figure 3. Trial-level association for the advanced colorectal cancer example. Circles (with size proportional to the sample size of the trial) represent the pairs of the estimated treatment effects on the surrogate and overall survival. The solid straight line is the linear regression providing the predicted treatment effect on overall survival. The squares correspond to the results of the two validation trials that investigated addition of irinotecan. The diamond corresponds to the results of the validation trial that investigated addition of oxaliplatin. Note that both axes are on a log scale.

References

[1] Buyse M, Molenberghs G, Burzykowski T, Renard D, Geys H (2000) The validation of surrogate endpoints in meta-analyses of randomized experiments. Biostatistics 1: 49–68.

[2] Buyse M, Burzykowski T, Carroll K, Michiels S, Sargent D, Miller LL, Elfring GL, Pignon JP, Piedbois P (2007) Progression-free survival is a surrogate for survival in advanced colorectal cancer. Journal of Clinical Oncology 25: 5218–5224.

[3] Buyse M (2016) Introduction. In: Applied Surrogate Endpoint Evaluation Methods with SAS and R (Alonso A, Bigirumurame T, Burzykowski T, Buyse M, Molenberghs G, Muchene L, Perualila NJ, Shkedy Z, Van der Elst W), Chapman & Hall/CRC (ISBN978-1-4822-4936-1).

[4] Buyse M, Saad ED, Burzykowski T, Regan MM, Sweeney CS (2022) Surrogacy beyond prognosis: the importance of „trial level" surrogacy. The Oncologist 5: 266 271.

[5] HTA CG (2024) https://health.ec.europa.eu/publications/guidance-outcomes-joint-clinical-assessments_en.

[6] Oba K, Paoletti X, Alberts S, Bang YJ, Benedetti J, Bleiberg H, Catalano P, Lordick F, Michiels S, Morita S, Ohashi Y, Pignon JP, Rougier P, Sasako M, Sakamoto J, Sargent D, Shitara K, Van Cutsem E, Buyse M, Burzykowski T on behalf of the GASTRIC group (2013) Disease-free survival as a surrogate for overall survival in adjuvant trials of gastric cancer: a meta-analysis. Journal of the National Cancer Institute 5: 1600–1607.

[7] Paoletti X, Oba K, Bang YJ, Bleiberg H, Boku N, Bouche O, Catalano P, Fuse N, Michiels S, Moehler M, Morita S, Ohashi Y, Ohtsu A, Roth A, Rougier P, Sakamoto J, Sargent D, Sasako M, Shitara K, Thuss-Patience E, Van Cutsem E, Burzykowski T, Buyse M on behalf of the GASTRIC group (2013) Progression-free survival as a surrogate for overall survival in patients with advanced/recurrent gastric cancer: a meta-analysis. Journal of the National Cancer Institute 5: 1608–1612.

[8] Shi Q, de Gramont A, Grothey A, Zalcberg J, Chibaudel B, Schmoll H-J, Seymour MT, Adams R, Saltz L, Goldberg RM, Punt CJA, Douillard J-Y, Hecht JR, Hurwitz H, Diaz-Rubio E, Proschen R, Tebbutt NC, Fuchs C, Souglakos J, Falcone A, Tournigand C, Kabbinavar FF, Heinemann V, Van Cutsem E, Bokemeyer C, Buyse M, Sargent DJ for the Analysis and Research in CAncers of the Digestive System (ARCAD) Group (2014) Individual patient data analysis of progression-free versus overall survival as a first-line endpoint for metastatic colorectal cancer in modern randomized trials: Findings from 16,700 patients from the ARCAD database. Journal of Clinical Oncology 33: 22 28.

What is a patient-relevant additional benefit? Seeking a structured discourse for a consensus process

By Florian Staeck

Even after almost 15 years of early benefit assessment since the introduction of AMNOG, there is still no full consensus among the key stakeholders on how a patient-relevant additional benefit can be defined in individual cases. As much as the paradigm shift that the AMNOG procedure gives high priority to the highest available evidence is generally welcomed, the determination of this evidence, particularly in relation to surrogate parameters, remains a contentious issue among stakeholders. This is especially the case when clinically relevant endpoints could have been investigated, but this opportunity was missed.

This issue was highlighted during the 20th meeting of the Interdisciplinary Platform for Benefit Assessment, which was held in Berlin on 27/28 September 2024 under the title „Which outcomes are relevant for patients?". At the conference, it was repeatedly emphasised that it would be desirable to have evidence-based procedures by which patient relevance could be determined by consensus.

However, it was stated that jurisprudence was unlikely to provide much clarification on the associated questions, according to statements made at the meeting. This is because patient relevance is an undefined legal term that is highly open to interpretation and lacks precise substantive definition. Within the AMNOG system, the Federal Joint Committee (G-BA) has been granted the relevant scope for judgement. Benefit assessments are, in this regard, acts of subordinate legal norm-setting. And they were only subject to judicial review to a very limited extent. This was generally restricted to determining, e.g. whether all relevant jurisdictional, procedural, and formal requirements have been met, or whether irrelevant considerations have taken place.

A G-BA decision would only be deemed unjustifiable if it contradicted the generally accepted state of HTA science. On the other hand, courts are not allowed to substitute their own assessments for those of the G-BA. As far as can be seen, no legal proceedings challenging a benefit assessment decision by the G-BA has ever aimed to establish an endpoint as patient-relevant. In other words: jurisprudence is passing the ball back to science and the players in the AMNOG cosmos on the disputed issues.

The same applies to politics, as was made clear during the meeting. It was stated that the German healthcare system was characterised by a delegation of central discussions on medical progress to self-governance bodies, while the legislature limited itself to defining the evaluation framework, such as through the Pharmaceutical Benefit Assessment Regulation (AM-NutzenV). This also applies to the controversial question of how a patient-relevant additional benefit should be defined. Legislative intervention would only be conceivable if it became apparent that medical progress was no longer reaching patients in the statutory health insurance system, it was stated.

Better representation of the chronification of diseases

If the self-administration itself was no longer able to find effective solutions, the legislator or regulator would be overwhelmed by the sheer volume of the debate. It was stated that the example of the United Kingdom illustrates the problems associated with a politicized decision-making culture in healthcare.

In Germany, it was explained that situations occasionally arise where surrogate parameters accepted in the context of pharmaceutical approval are not recognised to the same extent in early benefit assessment. However, the task of developing common standards for the assessment of surrogate parameters is becoming increasingly complex due to medical progress – rather than simpler. This was particularly relevant given that diseases that were

previously often fatal now sometimes follow a chronic course. Some participants therefore argued that HTA criteria should be adjusted to better reflect the chronic nature of certain diseases.

One aspect that must be avoided at all costs, it was argued, was the emergence of „parallel worlds" in pharmaceutical assessment, driven e.g. by the chronic nature of oncological diseases. Instead of overall survival (OS) as the primary endpoint, composite endpoints should be used – depending on the medical field – to measure not only disease progression but also quality of life. Participants questioned whether an indication-specific weighting of endpoints could be a practicable strategy in this regard.

This view was particularly supported by those participants who argued that rigidly predefined patient-relevant endpoints should be reconsidered, especially in cases where treatment response varies significantly between individuals. It was argued that a „one-size-fits-all" approach was deemed inappropriate, as patient relevance is not a static variable. This was particularly true in the field of psychiatry, where patient-reported outcomes (PROs) were central outcomes. Only PROs could provide insight into how a patient experiences their disease and whether the therapeutic intervention could provide a patient-relevant benefit.

It was noted that disease-related functional impairments were perceived very differently by patients. In psychiatry, a speciality in which biomarkers only exist in exceptional cases, PROs remain the key outcomes. Mortality was generally not suitable for capturing treatment results in this area. In many cases, valid surrogate markers might include whether the patient was in paid employment and whether the patient was in a stable relationship. The patient-centred outcome should therefore be a relapse-free and functional long-term survival of the patient.

The meeting also highlighted divergent views as to why little or no progress has been made in the interpretation of surrogate endpoints in Germany in recent years. This is also due to the lack of a structured debate, it was said. From IQWiG's point of view, only in a few cases has a surrogate endpoint been validated satisfactorily. Validations were only submitted very rarely. In addition, for around two-thirds of benefit assessments on chronic diseases, suitable clinical trial data were lacking – a situation that had remained unchanged for years. Although many generic and disease-specific scores had been established, these must also be appropriately incorporated into clinical studies – an area where there is still room for improvement.

Discussion about response thresholds

Participants also disagreed on the interpretation of response thresholds in responder analyses. These analyses assess whether the proportion of patients experiencing a noticeable change in an endpoint differs between the two treatment groups in a study. IQWiG's position is that from a response threshold value of 15%, a small but sufficiently certain noticeable change can be assumed. This value had proven to be practicable since its introduction. Moreover, fundamental issues with the use of Minimal Important Differences (MID) remained unresolved. It was reported that developments in MID research were still being monitored.

Focus on the evaluation relevance of PRO data

A counterargument was raised that the 15% response threshold constituted a rigid German „universal formula", which failed to take into account specific therapeutic situations and did not adequately consider the problem of sufficient study power. Furthermore, there was no systematically collected quantitative evidence to support this thres-

hold. Since 2014, the proportion of benefit assessment procedures in which data on quality of life is submitted had been around 70%. PRO data were thus collected but had not yet achieved a level of significance in the data hierarchy making them truly relevant for determining additional benefit, was the criticism.

Other participants were convinced that a „learning curve" had been observed in the selection of suitable study endpoints and comparator therapies in recent years – this was partly due to the fact that the Act for Greater Safety in the Provision of Medicines (GSAV) created the possibility of involving medical associations in consultations on the selection of a suitable comparator therapy at an early stage. The number of corresponding requests for advice from the G-BA recently totalled around 180 per year, they said.

In fact, it became clear at the meeting that the validation of surrogate parameters would remain a highly challenging methodological task in the future. If clinically relevant symptoms could not be observed in a study, a methodology must be available to predict an individual relationship between the surrogate and the outcome.

It was explained that the prerequisite for such predictions was a stable test environment, which was often not the case for most diseases.

To make matters worse, even if a surrogate proved to be valid in the context of a specific treatment, this does not necessarily apply to another treatment. The variance in the prediction remained so high in many cases that the validation of a surrogate was not reliably possible even if data from large patient cohorts were used, it was explained.

However, surrogates such as HBA1c in type 1 diabetes, eGFR, or creatinine clearance in renal failure or Sustained Virological Response (SVR) in viral diseases had now been accepted as clinically relevant surrogates.

Finally, different positions emerged at the meeting regarding the significance of the European benefit assessment (EU HTA), which was due to start in January 2025. On the one hand, it was pointed out that the definition of relevant outcomes in the scoping procedure would continue to be the responsibility of the member states in the EU HTA. This means that there would be no ranking or categorisation of outcomes as part of the European assessment. All endpoints in the future Joint Clinical Assessments (JCA) would thus be considered equal.

The corresponding guidance documents for JCA stipulate that validated surrogate endpoints can be used in addition to patient-relevant endpoints if a member state considers this to be relevant. Against this background, it was suggested that the previously heterogeneous national discourses on patient-relevant outcomes were unlikely to be harmonised.

In contrast, doubts were raised as to whether the JCAs would remain completely value-free and solely „technical" reports. Rather, it was expected that there would be a gradual convergence of previously differing national assessment standards.

ADVISORY BOARD OF THE INTERDISCIPLINARY PLATFORM ON BENEFIT ASSESSMENT

Dr Jürgen Bausch
Britta Bickel
Wolfgang van den Bergh
Prof Dr Wolfgang Greiner
Dr Ulrike Götting
Dr Antje Haas
Dr Harald Herholz
Dr Ulf Maywald
Dr Hendrik Pugge
Dr Daniela Preukschat
Dr Heinz Riederer
Prof Dr Jörg Ruof
Prof Dr Bernhard Wörmann

DISCUSSANTS

Dr Jürgen Bausch
Britta Bickel
Prof Dr Karl Broich
Prof Dr Thomasz Burzykowski
Claus Burghardt
Dr Jan Daniels-Trautner
Dr Martin Danner
Prof Dr Peter Falkai
Dr Mathias Flume
Edith Frénoy
Dr Ulrike Götting
Dr Rosa Gonzalez-Quevedo
Dr Harald Herholz
Antje Hoppe
Sabine Jablonka
Dr Florian Jantschak
Dr Elaine Julian
Prof Dr Sebastian Kluckert
Dr Werner Kulp
Dr Norbert Marschner
Dr Ulf Maywald
Sabine Mentrup
Anja Moeller
Dr Paula Piechotta
Dr Daniela Preukschat
Dr Hendrik Pugge
Dr Andrey Rasch
Dr Heinz Riederer
Prof Dr Jörg Ruof
Dr Christian Rybak
Dr Felix Schönfeldt
Stefan Schwartze
Dr Johanna Seeger
Eva Stumpe
Wolfgang van den Bergh

IMPRINT

PUBLISHER
Editorial Advisory Board of the Interdisciplinary Platform:
Dr Harald Herholz
Kassenärztliche Vereinigung Hessen
Europa-Allee 90
60486 Frankfurt
Germany

PUBLISHING COMPANY
Springer Medizin Verlag GmbH
Am Forsthaus Gravenbruch 5
63263 Neu-Isenburg, Germany
German Commercial Register (HRB):
Berlin Charlottenburg District Court
HRB: 167094 B
VAT-ID: DE 230026696

Telephone: +49 6102 5060
Mail address: info@aerztezeitung.de

EDITORIAL WORK
Dr Florian Staeck

AUTHORS
Stefan Schwartze
Dr Martin Danner
Eva Stumpe
Prof Dr Sebastian Kluckert
Dr Jutta Wendel-Schrief
Prof Dr Bernhard Wörmann
Dr Daniela Preukschat
Dr Sebastian Meller
Dr Andreas Rhode
Prof Dr Peter Falkai
Prof Dr Tomasz Burzykowski
Dr Andrej Rasch
Dr Johanna Seeger
Prof Dr Jörg Ruof

PICTURE CREDITS
The Little Hut / stock.adobe.com

LAYOUT/GRAPHICS
Sandra Bahr
Oliver Hippmann

PRINT
F&W Druck- und Mediencenter GmbH
Holzhauser Feld 2, 83361 Kienberg

Springer Medizin

In cooperation with and with the kind support of AbbVie Deutschland Gmbh & Co. KG, DAK Gesundheit, MSD Sharp & Dohme GmbH, Novo Nordisk Pharma GmbH, Roche Pharma AG, Verband Forschender Arzneimittelhersteller e.V., Xcenda

References to the speaker portraits

Stefan Schwartze: Jan Pauls | Martin Danner: BAG Selbsthilfe | Eva Stumpe: Privat | Bernhard Wörmann: DGHO | Sebastian Kluckert: Ralf Baumgarten| Jutta Wendel-Schrief: MSD | Andreas Rhode: Olaf Plotke/MDWL | Peter Falkei: Fotostudio Sauter | Jörg Ruof: European Access Academy | Thomasz Burzykowski: International Drug Development Institute | Andrej Rasch: vfa / B. Brundert | Johanna Seeger: Studioline Photography

INTERDISCIPLINARY PLATFORM ON BENEFIT ASSESSMENT

Which endpoints are patient-relevant?

Volume 20
March 2025