GUIDELINE

Endpoints used in relative effectiveness assessment of pharmaceuticals

Surrogate Endpoints

Final version

February 2013

The primary objective of EUnetHTA JA1 WP5 methodology guidelines is to focus on methodological challenges that are encountered by HTA assessors while performing a rapid relative effectiveness assessment of pharmaceuticals.

The guideline “Endpoints used in REA of pharmaceuticals: surrogate endpoints” has been elaborated by experts from NOKC and HAS, reviewed and validated by all members of WP5 of the EUnetHTA network; the whole process was coordinated by HAS. As such the guideline represents a consolidated view of non-binding recommendations of EUnetHTA network members and in no case an official opinion of the participating institutions or individuals.

Table of contents

Acronyms – Abbreviations

AHRQ	Agency for Healthcare Research and Quality
CHMP	Committee of Human Medicinal Products
EFPIA	European Federation of the Pharmaceutical Industries and Associations
EMA	European Medicines Agency
HTA	Health Technology Assessment
IOM	Institute of Medicine
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
MA	Marketing Authorization
RCT	Randomized Controlled Trial
REA	Relative Effectiveness Assessment
STE	Surrogate Threshold Effect
HPV	Human Papilloma Virus
CIN	Cervical Intraepithelial Neoplasia

Summary and Recommendations

Summary

Surrogate endpoints act as substitutes for clinical endpoints and are expected to predict the effect of therapy (benefit and/or harm). An improvement in surrogate endpoint may be or may not be perceived by the patient. In many cases, surrogate endpoints do not themselves directly measure a clinical benefit.

A biomarker can be used as a surrogate endpoint if it acts as a substitute for a clinical endpoint that directly measures clinical benefit. When attempting to validate a biomarker as a surrogate endpoint reliably predicting a final clinical endpoint, the evaluation process should consider the following three steps: analytical validation based on extensive documentation, qualification and utilization (IOM 2011).

For the purpose of REA, both biomarkers and intermediate endpoints will be considered if used as surrogate endpoints to substitute for a clinical (final) endpoint.

The acceptability of an endpoint as a surrogate endpoint for a specific clinical endpoint is based on its biological plausibility and empirical evidence. However, there is still a need to determine which surrogate endpoints can reliably predict clinical benefit relevant for relative effectiveness assessment (REA) purposes. Hence, more scientific and clinical knowledge is required for the validation of surrogate endpoints in both drug development and relative effectiveness assessment. In this review we have considered when it is appropriate to use surrogate endpoints and what conditions they have to fulfil for the purpose of REA.

A literature search was performed in databases and websites for reviews, policy papers and methodological guidelines describing when it is appropriate to use surrogate endpoints for relative effectiveness assessment. In addition, experience from recent reimbursement decisions based on surrogate endpoints was also summarised and taken into account while giving recommendations.

The use of surrogate endpoints in the assessment of (added) clinical benefit of a health technology is controversial, since the validity of surrogate endpoints has rarely been rigorously fully established; only few surrogate endpoints have been shown to be true measures of tangible clinical benefit.

Final clinical endpoints are preferred both for first assessment and re-assessment in the performance of REA of pharmaceuticals. However, for the **initial (first) assessment**, surrogate endpoints might be accepted if the validity of the surrogate/final clinical endpoint relationship has been previously clearly established for clinical endpoints of interest for REA. The availability of a large safety database for a pharmaceutical that has shown effectiveness only on a surrogate endpoint is particularly important.

For **the re-assessment**, in principle, effectiveness should be demonstrated whenever possible on final clinical morbidity and mortality endpoints (e.g. stroke, myocardial infarction, fracture).

Comparative clinical data (or evidence that data will be provided in a reasonable timeframe) on final/clinical endpoints coming from post-marketing clinical trials and/or other sources should be provided before the re-assessment of a pharmaceutical.

Recommendations

Recommendation 1 The REA of pharmaceuticals should be based whenever possible on final patient-relevant clinical endpoints (e.g. morbidity, overall mortality).
Recommendation 2 In the absence of evidence on final patient-relevant clinical endpoint that directly measures clinical benefit, both biomarkers and intermediate endpoints will be considered as surrogate endpoints in REA if they can reliably substitute for a clinical endpoint and predict its clinical benefit.
Recommendation 3 If surrogate endpoints are used for REA, they should be adequately validated: the surrogate-final endpoint relationship must have been demonstrated based on biological plausibility and empirical evidence. The level of evidence, the uncertainties associated and the limits of their use should be explicitly explained. Complete validation data should always be provided. For adequately validated surrogate endpoints, a second validation for REA purposes will not be necessary.
Recommendation 4 Validation of a surrogate versus patient-relevant clinical endpoint is normally undertaken in a specific population and for a specific drug intervention i.e. validation is disease-specific, population-specific and pharmaceutical class (technology) specific. Demonstration of surrogate validation both within and across drug classes should be thoroughly justified.
Recommendation 5 For the **first assessment,** even if final endpoints are preferred, surrogate endpoints might be accepted if the validity of the surrogate/final clinical endpoint relationship has been previously clearly established on clinical endpoints of interest for REA. The availability of a sufficiently large safety database is particularly important. Evidence on safety outcomes should always be reported.
Recommendation 6 For the **re-assessment**, effectiveness should in principle be demonstrated on morbidity and mortality endpoints (e.g. stroke, myocardial infarction, fracture). Comparative clinical data (or evidence that data will be provided in a reasonable timeframe) on relevant clinical endpoints and safety coming from post-marketing clinical trials and other sources should be provided whenever possible before the re-assessment of the pharmaceutical can be carried out. The absence of data on clinical endpoints relevant for REA might be acceptable when a clinical endpoint is difficult or impossible to study (very rare or delayed) or target population is too small to obtain meaningful results on relevant clinical endpoints even after very long follow-up (very slowly progressive and/or rare diseases). However, these exceptions need to be carefully argued and agreed in advance of an REA.

Recommendation 7 Re-assessment requirements for further data regarding relevant clinical endpoints should be clearly defined when a REA has been previously made based on surrogate endpoints for the first assessment.
Recommendation 8 Further methodological research on the use of surrogate outcomes is needed to inform future REA approaches for the handling of surrogates.

1. Introduction

1.1. Definitions and general information

1.1.1. Definitions

Surrogate endpoint

A clinical endpoint is a "direct measure of how a patient feels, functions or survives" (Biomarkers Definitions Working Group 2001).

A surrogate endpoint is an endpoint that is intended to replace a clinical endpoint of interest that cannot be observed in a trial - it is a variable that provides an indirect measurement of effect in situations where direct measurement of clinical effect is not feasible or practical. (ICH Guideline E9, Statistical Principles for Clinical Trials, 1998).

A surrogate endpoint may be a biomarker that is intended to substitute for a clinical endpoint. A surrogate endpoint may also be a clinical endpoint that is used to replace the endpoint of interest, such as an intermediate clinical endpoint.

A surrogate endpoint is expected to predict clinical benefit (or harm or lack of benefit or harm) based on epidemiologic, therapeutic, pathophysiologic, or other scientific evidence. (Biomarkers Definitions working group 2001). In many cases, however, an effect on a surrogate endpoint will not per se be of any benefit to the patient (biomarkers are typical examples).

Biomarker

A biomarker can be defined as a characteristic that is objectively (reliably and accurately) measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to an intervention (Biomarkers Definitions Working group 2001). Example: cholesterol level, HbA1c.

The biomarker must lie on the pathophysiologic causal pathway of the disease; it must be correlated with a clinical endpoint to be useful in detecting disease and assessing prognosis, and validated through a validation or qualification process (Fleming et al. 1996).

A surrogate endpoint represents a special use of a biomarker, in which the biomarker substitutes for a clinical endpoint.

Intermediate endpoint

An intermediate endpoint is a clinical endpoint such as measure of a function or of a symptom (disease-free survival, angina frequency, exercise tolerance) but is not the ultimate endpoint of the disease, such as survival or the rate of irreversible morbid events (stroke, myocardial infarction) (Temple et al. 1999).

Improvement in an intermediate endpoint due to treatment is well perceived and can be of value to the patient even if it does not lead to the improvement of morbidity or mortality.

For the purpose of REA, both biomarkers and intermediate endpoints will be considered, if used, as surrogate endpoints to substitute for a clinically meaningful (final) endpoint.

1.1.2. General information

The possible use of surrogate endpoints includes different situations.

For drug registration purposes, a biomarker that can accurately predict and quantify clinical benefit or harm, can be used as a surrogate for clinical outcome. Before a biomarker can be accepted as a surrogate endpoint, there is a need to have confidence that changes in the biomarker reliably predict changes in the desired clinical endpoints (EMA 2007). In addition, during drug development, there is a possibility to use surrogate endpoints in order to assess biological activity of a new medicine in phase II trials and to make a go – no go decision for a phase III trial and/or to select an adequate subpopulation in drug development.

In the context of drug approval, surrogate endpoints are used if clinical events are rare/delayed (very slowly progressive diseases, very long follow-up needed for their assessment, and rare diseases) and/or life-threatening diseases with no therapeutic alternative.

In the context of REA, surrogate endpoints are often assessed to demonstrate relative effectiveness of pharmaceuticals both at first assessment and at re-assessment, as they can be relatively quickly and easily measured and with high precision. They are often used to replace a distal endpoint with a more proximal one that can be measured earlier (in situations where assessment of final clinical endpoint requires long follow-up) (Cochrane handbook)) in order to spare time and money and allow for rapid decision making.

The use of surrogate endpoints in the assessment of the benefit of a health technology is controversial, since the validity of surrogate endpoints has rarely been rigorously established in this context; only few surrogate endpoints have been shown to be true measures of tangible clinical benefit (Fleming et al. 2005) of interest for REA.

1.2. Context

1.2.1. Problem statement

Are results on surrogate endpoints acceptable for relative effectiveness assessment (REA)? Which type of surrogate endpoints are used in a REA? What do they have to fulfill? When is it justified to use them?

1.2.2. Discussion (on the problem statement)

Surrogate endpoints are often used in clinical trials for drug development (EMA 2007). They are also frequently assessed for REA of pharmaceuticals. However, there is still a need to decide which surrogate endpoints can reliably predict clinical endpoints and therefore be acceptable for the purpose of REA. In addition, more fundamental scientific and clinical knowledge is needed for the qualification of biomarkers as surrogate endpoints as well as for acceptability of surrogate endpoints for REA in order to support decisions.

1.3. Scope/Objective(s) of the guideline

This guideline is intended to provide guidance on when and how surrogate endpoints can be used for REA of pharmaceuticals.

The guideline is not intended to give a comprehensive list of validated surrogate endpoints and how well they predict final clinical endpoints.

In addition, the following is out of scope of this guideline:
- The description of the process of validation of biomarkers: currently, there is no systematic, transparent and widely agreed-upon process of biomarker validation; a list of validated biomarkers does not exist. Therefore, only principles of validation will be mentioned to help HTA assessors. However, the guideline fully covers the situation when a biomarker is used as a surrogate endpoint.

- The use of surrogate endpoints for diagnostic or screening purposes: the guideline concentrates on the use of surrogate endpoints to replace final clinical endpoints when performing REA of pharmaceuticals

- A detailed review of statistical methods of the validation of surrogate endpoints: principles of statistical reasoning as well as relevant literature sources will be mentioned.

1.4. Relevant EunetHTA documents

This guideline should be read in conjunction with the following documents:

EUnetHTA guideline on Endpoints used in REA of pharmaceuticals: Clinical Endpoints

2. Summary of the analysed literature

Validation and use of surrogate endpoints for efficacy/effectiveness assessment

Validation of surrogate endpoints

A surrogate endpoint may be considered validated if it is sensible, measurable, interpretable and highly accurate in predicting the clinically relevant endpoint. A surrogate endpoint is correlated to the final clinical endpoint if it fully captures the net effect of intervention on all mechanisms that influence the clinical outcome and reflect the totality of the effect. In addition, the intervention on the surrogate endpoint must predict the effect on the clinical endpoint (Fleming 1996).

When attempting to validate a biomarker as a surrogate endpoint reliably predicting a final clinical endpoint, the evaluation process may be generally based on the following three steps (IOM 2011):

- Analytical validation: is the biomarker able to be accurately (accuracy: reliability, reproducibility, sensitivity and specificity) measured?
- Qualification: is the biomarker invariably associated with the clinical endpoint of interest? Does the intervention of interest affect both biomarker and the clinical endpoint in the same way?
- Utilization: what is the context of the proposed use of a biomarker? (a substitute for the effect of one pharmaceutical class? Of different pharmaceutical classes given for the same indication? For a disease state?)

The acceptability of a surrogate endpoint in supporting effectiveness of a pharmaceutical is mostly based on its:

- **biological plausibility**: evidence that the surrogate endpoint is on the causal pathway to the clinical outcome (preferably close to the outcome), credible animal models, well understood pathogenesis of a disease, and well understood mode of action of the proposed drug

and

- **empirical evidence**:

    - **results from randomised clinical trials**: comparative treatment effect of the proposed drug versus best therapeutic alternative on a surrogate that has satisfactorily predicted a true clinical outcome. High correlation of effects on surrogate and on clinical endpoint of interest must be demonstrated (so that an association of the form "the larger the effect on the surrogate, the larger the effect on clinical endpoint" can be reasoned). Generalization to other drugs, even of the same pharmacologic class, and also generalization regarding other indications or areas of indications has to be checked carefully.

    - **observational evidence** relating to the association between the surrogate and clinical outcome of interest independently of treatment effects.

The acceptability of a surrogate endpoint has also been based on **other risk-benefit and/or public health** considerations such as a serious life-threatening disease with no alternative therapy, large safety database available, difficult to study clinical endpoint (very rare or delayed).

Establishing that a surrogate lies on a causal pathway and is correlated with a clinical outcome is important but not sufficient to validate a surrogate. In addition, it should be demonstrated that modification of a surrogate without and with therapeutic intervention reliably modifies the clinical outcome (Aussagekraft von Surrogatendpunkten in der Onkologie, IQWIG 2011).
The evidence for the validation of the surrogate-final outcome relationship has been presented by taking into account the level of evidence:

- level 1: evidence demonstrating that treatment effects on the surrogate endpoint correspond to effects on the patient-related clinical outcome (from clinical trials); comprises a meta-analysis of several RCT and establishment of correlation between effects on the surrogate and clinical endpoint
- level 2: evidence demonstrating a consistent association between surrogate endpoint and final patient-related endpoint (from epidemiological/observational studies); and
- level 3: only evidence of biological plausibility of relationship between surrogate endpoint and final patient-related endpoint (from pathophysiological studies and/or understanding of the disease process).

Additionally, a hierarchy for endpoints, depending on the levels of evidence available, has also been proposed for drug registration purposes (Fleming 2005):
- true clinical efficacy endpoints as level 1;
- validated surrogate endpoints as level 2;
- non validated surrogate endpoints that are "reasonably likely to predict clinical benefit" as level 3;
- correlates that are solely a measure of biological activity as level 4

The validation of surrogates appears to be:

- **disease-specific** (stage of disease-specific): validation is performed within an indication; validity of a surrogate should be demonstrated for different stages of a disease

- **population-specific**: validity if a surrogate should be justified for different patient populations (age, gender, co-morbidities) with a disease.

- **pharmaceutical class (technology) specific**: a surrogate (e.g. LDL-C) has to be validated for each pharmaceutical class of drugs separately (LDL-C validated for statins but not for fibrates).

A surrogate may then be validated for several pharmaceutical classes within an indication; an attempt to extrapolate the validity of a surrogate to other pharmaceutical classes within an indication is always difficult and has to be thoroughly justified.

Here are some examples of surrogate endpoints that have been used for assessment both in the context of drug approval and REA purposes, such as: HbA1c as a measure of glycaemic control; glycaemic control as a surrogate for the avoidance of long-term complications in patients with diabetes; a decrease in serum cholesterol levels (LDL-C) induced by HMG-CoA reductase inhibitors as a surrogate for a decrease of cardiovascular morbidity and mortality.

In addition to efficacy/effectiveness assessment based on surrogate endpoints, it is important to take the safety measures into account while assessing a pharmaceutical both for marketing authorisation and for REA. Even if a surrogacy has been demonstrated on a specific efficacy endpoint, unexpected side effects of a pharmaceutical may lead to an increased mortality or any other unfavourable outcome. Therefore, such an increase of rate of side effects is important to be assessed both pre- and post-marketing, in large randomised trials and through prospective post-marketing surveillance data.

Overviews of statistical methods for the validation of surrogates have been given (see Bibliography – Statistical methods). The majority of the procedures, even those that have been applied to real data examples, rely on meta-analyses of several RCTs and estimate the correlation of the effects on the surrogate and the effects on the clinical endpoint. There is no clear consensus of which correlation values are sufficient to assume adequate surrogacy, but values of between about 0,85 and 0,95 are often discussed. If there is no high correlation demonstrated, conclusions might still be made if the surrogate threshold effect (STE) is considered (Burzykowski et al., 2005). Also based on an analysis of several RCTs, the STE defines the minimum absolute value of the effect on the surrogate which has to be observed in a new trial to deduce an effect on the clinical endpoint. Accordingly, the STE can be computed for a certain level of change in a biomarker that will translate into clinical benefit. In both cases, certainty of the conclusions depends on pre-specified levels of significance.

A detailed systematic review of statistical methods for the validation of surrogates is out of scope of this guideline.

3. Discussion and conclusion

The objective of this guideline is to give methodological guidance on the use and acceptability of surrogate endpoints in a REA of pharmaceuticals. Whether a surrogate endpoint is acceptable for REA is a matter of scientific judgement; there is no recommendation that can be universally accepted.

The aim of the considerations given here is to indicate the amount of information necessary to minimise uncertainty while judging the acceptability of a surrogate for the purpose of REA; it also tries to explain the underlying rationale. When a relationship between the surrogate and final clinical endpoint has been established, the surrogate can be used to predict the expected benefit on the final clinical endpoint relevant for REA. However, it is important when doing this to incorporate all sources of uncertainty into this prediction, including uncertainty around the relationship.

The literature analysis indicates a cautious approach of HTA institutions to the use of surrogate endpoints in health technology assessment. Whenever possible, the REA of health technologies should rely on clinical endpoints; if surrogate endpoints are to be used, their acceptability for the purpose of REA and their results should be interpreted with caution (Velasco Garrido et al., 2009). Different actors (physicians and policymakers) involved in health system decision-making have to deal with the issue of what is a valid parameter to assess health benefit. This should be dealt with carefully while assessing relative effectiveness of pharmaceuticals.

For REA of pharmaceuticals, surrogate endpoints expected to predict the effect of therapy on relevant clinical endpoints have been assessed in many cases. However, there may be differences in the endpoints considered relevant by HTA institutions/reimbursement organisations and regulatory authorities. In addition, there may be differences in HTA requirements between the **first assessment** of a pharmaceutical rapidly after marketing authorisation has been granted and the **re-assessment** of a pharmaceutical or the entire pharmaceutical class within an indication.

Final clinical endpoints are preferred both for first assessment and re-assessment in the performance of REA of pharmaceuticals. However, in the context of first assessment, shortly after Marketing Authorization (MA) has been granted (this is sometimes done in parallel), even if final clinical endpoints are preferred, surrogate endpoints might be accepted for assessment if the validity of the surrogate/final REA-relevant clinical endpoint relationship has been previously clearly established and data on all validation steps provided. In addition, for a pharmaceutical that has shown effectiveness only on a surrogate endpoint, the availability of a large safety database is especially important.

Reimbursement decisions are necessarily based on several factors (endpoints other than surrogates, harms and related uncertainties), and may vary from one institution to another, from one pharmaceutical to another and for different levels of proof provided (e.g. reimbursement decisions may differ for a statin that has shown effectiveness on final clinical endpoints and a statin that has shown effectiveness on LDL-C only). Relevant clinical endpoints (e.g. mortality, morbidity events such as stroke, myocardial infarction, heart failure) are influenced by several patho-physiological mechanisms (the effect on the surrogate being only one of them). Therefore, for the re-assessment, effectiveness should be demonstrated whenever possible on final clinically relevant morbidity and mortality endpoints (e.g. stroke, myocardial infarction, fracture). Companies are recommended to provide comparative clinical data on final/clinically relevant endpoints coming from post-marketing clinical trials and observational data sources before the re-assessment of their pharmaceutical. It is important to have evidence that these post-marketing commitments will be fulfilled within a reasonable time-frame. The absence of such evidence might directly influence (level of) reimbursement decisions.

There might be situations where REA based on surrogate endpoints could be acceptable in case of very slowly progressive and rare diseases even at re-assessment, such as when the clinical event happens after a very long time or when the clinical event intended to be avoided is very rare and difficult to obtain in the context of clinical trials. The same is true when a target population is too small to obtain meaningful results on relevant clinical endpoints even after a long period of time (e.g. mortality in milder stages of pulmonary arterial hypertension). In such situations, surrogate endpoints could be acceptable if they reliably predict rare and late clinical events.

In the field of oncology and haematology, especially when there is a high unmet need, and/or no available therapeutic alternative, the use of intermediate endpoints such as progression-free survival (PFS) appears acceptable for accelerated/conditional MA approval. In the REA setting, the acceptability of intermediate endpoints for these drugs still vary in different EU countries, and are done on the case by case basis. In oncology, PFS is an intermediate endpoint that is relevant on its own right. However, data only on PFS without data on OS or at least on HRQoL or other patient-relevant endpoints is considered insufficient for REA. In addition, the point of time at which the REA is done is important; decisions may vary depending on the maturity of data on final endpoints, for example, if assessment is done when there is no or insufficient information on clinically relevant endpoint as compared to situations where there is data on both surrogate and final endpoint. The acceptability of progression free survival has not the same impact in adjuvant or metastatic disease. In the adjuvant setting, PFS appears acceptable; in the metastatic setting, PFS alone is insufficient; it might be considered if coupled with quality of life assessment and survival data, the maturity of which will be considered on the case by case basis.

The reasons for differences in assessments may be related to country-specific factors (such as local context and values), but also to different interpretation of data coming from the same primary studies (Kristensen 2010). The recent assessments of two human papilloma virus (HPV) vaccines from six agencies in Europe point out several differences in conclusions that may have been related to different interpretations of different surrogate measures (proportion of vaccine HPV types-associated pre-cancerous states as a surrogate for cancers avoided, etc). For this reason, a surrogate which is a close and necessary step in the development of clinical outcome should be preferred to earlier endpoints (such as CIN3 for the development of cervical cancer as compared to CIN1).

How reliable is a surrogate endpoint for final patient benefit and/or absence of harm assessment? Recently we have learned a lot from the withdrawal of rosiglitazone. When measuring HbA1c lowering effect (as a surrogate for glycaemic control), rosiglitazone appears to perform as well as the other oral antidiabetics, however when measuring relevant clinical outcomes, rosiglitazone performs worse than other drugs from the same pharmaceutical class; therefore, caution has been recommended while assuming class effect as well as the necessity to assess safety of a new pharmaceutical as thoroughly as possible (IOM 2011).

Annexe 1. Bibliography

References used for search:

1. Biomarkers Definitions Working Group.Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin Pharmacol Ther. 2001 Mar;69 (3):89-95.

2. Bucher HC. Studien mit Surrogatendpunkten: Nutzen und Grenzen in der klinishen Entscheidungsfindung. Internist 2008; 49(6): 681-7.

3. Fleming TR and De Mets DL. Surrogate End Points in Clinical Trials: Are We Being Misled? Ann Intern Med. 1996;125:605-613.

4. Fleming TR. Surrogate endpoints and FDAs accelerated approval process. Health Affairs, 24, no.1, 2005:67-78.

5. Helfand M, Balshem H. AHRQ series paper 2: principles for developing guidance: AHRQ and the effective health-care program. J Clin Epidemiol. 2010 May;63(5):484-90.

6. Mangiapane S and Velasco Garrido M, Use of surrogate endpoints in HTA. GMS health technology assessment 2009, Vol. 5. Taylor RS, Elston J. The use of surrogate outcomes in model-based cost-effectiveness analyses: a survey of UK Health Technology Assessment reports. Health Technology Assessment, 2009;1-72.

7. Taylor RS, Elston J. The use of surrogate outcomes in model-based cost-effectiveness analyses: a survey of UK Health Technology Assessment reports. Health Technology Assessment, 2009;1-72.

8. Taylor RS and Elston J. The use of surrogate outcomes in model-based cost effectiveness analyses: a survey of UK Health Technology Assessment reports. Health Technology Assessment 2009; Vol. 13: No. 8.

9. Temple R. Are surrogate markers adequate to assess cardiovascular disease drugs? JAMA 1999; Vol 282: (No 8): 790-95.

10. Velasco Garrido M & Mangiapane S, Surrogate outcomes in health technology assessment: An international comparison. IJTAHC, 25:3 (2009), 315-22.

11. Ciani O, Buyse M, Garside R, Pavey T, Stein K, Sterne J, Taylor R. Comparison of treatment effect sizes associated with surrogate and final patient relevant outcomes in randomised controlled trials: meta-epidemiological study. *BMJ* 2013;346:f457

References for statistical methods:

12. Burzykowski T, Molenberghs G, Buyse M (Ed). The evaluation of surrogate endpoints. New York: Springer; 2005.

13. Aussagekraft von Surrogatendpunkten in der Onkologie, IQWiG report, 2011

14. Lassere MN. The Biomarker-Surrogacy Evaluation Schema: a review of the biomarker-surrogate literature and a proposal for a criterion-based, quantitative, multidimensional

hierarchical level of evidence schema for evaluating the status of biomarkers as surrogate endpoints. Stat Methods Med Res 2008; 17(3): 303-340.

15. Molenberghs G, Burzykowski T, Alonso A, Assam P, Tilahun A, Buyse M. A unified framework for the evaluation of surrogate endpoints in mental-health clinical trials. Stat Methods Med Res 2010; 19(3): 205-236.

16. Weir CJ, Walley RJ. Statistical evaluation of biomarkers as surrogate endpoints: a literature review. Stat Med 2006; 25(2): 183-203.

Guidelines:

17. Allgemeine Methoden Entwurf für Version 4.0 vom 15.11.2007 Kontakt: Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) Dillenburger Straße 27 D-51105 Köln E-Mail: methoden@iqwig.de https://www.iqwig.de/download/General_Methods_4-0.pdf

18. Cochrane Handbook

19. Report on the Surrogate to Final Outcome Working Group to the Pharmaceutical Benefits Advisory Committee

Additional references used:

20. Innovative drug development approaches. Final report from the EMEA/CHMP-think-tank group on innovative drug development. EMEA/127318/2007.

21. IOM (Institute of Medicine). 2011. Perspectives on biomarker and surrogate endpoint evaluation: Discussion forum summary. Washington, DC: The National Academies Press.

22. Report on the EMEA/CHMP Biomarkers Workshop. EMEA/522496/2006.

Annexe 2. Methods and results of literature search

A literature review has been conducted with the aim of locating studies that provide recommendations of when surrogate endpoints can be used and accepted for REA purposes.

Keywords for search

Keywords that were used for the bibliographic research:

- Biological marker
- Tumor marker
- Relative efficacy assessment
- Technology assessment, biological
- Evidence-based medicine
- Medical technology
- Treatment outcome
- Surrogate outcome
- Surrogate endpoint
- Surrogate marker
- Intermediary outcome

Sources of information

Data-bases

MEDLINE via OVID
Embase
Cochrane Database of Systematic Reviews
NHS Evidence
DARE

Websites

EMA
FDA
CCBAR
AHRQ
OSHU
NICE
CADTH
IQWIG
PBAC

Other

Google Scholar
The Cochrane Library
The Cochrane Methodology Register
Hand searches of references cited in relevant documents

Bibliographic search strategy

Where time limits could be specified the databases were searched for the period 01/01/1995 to 05/10/2010. The database searches were restricted to human subjects.

Database searches used the following search strategy:

[(biological marker OR surrogate marker OR surrogate endpoint OR surrogate outcome) AND (technology assessment, biological OR evidence-based medicine OR medical technology OR relative effectiveness assessment) AND (treatment outcome)]

Selection criteria

Publications were assessed on the basis of whether they discussed the use of surrogate markers or surrogate endpoints for the assessment of relative efficacy or relative effectiveness of health technologies. Review articles and methodological guidance documents assessing when it is appropriate to use surrogate markers and surrogate endpoints in health technology assessment or for relative effectiveness assessments were included in this review.

Documents that only applied surrogate endpoints for specific illnesses or conditions were excluded.

Annexe 3. Analysis and synthesis of the literature

Results of literature search

Summary of literature findings

Reference	Type of study	Aim of the study	Results and conclusions
Articles assessing the use of surrogate endpoints in REA, HTA, coverage decisions and treatment decisions			
RS Taylor and J Elston. The use of surrogate outcomes in model-based cost effectiveness analyses: a survey of UK Health Technology Assessment reports. Health Technology Assessment 2009; Vol. 13: No. 8.	Systematic Review	The study aimed to explore the use of surrogate outcomes in cost-effectiveness models within UK HTA Programme reports and provide a basis for guidance for their future use, validation and reporting.	Systematic review of 35 HTA reports included. 4 reports included modelling with surrogate endpoints. Recommendation 1: Ideally use final patient-related outcomes (i.e. mortality, important clinical events and HRQoL) Recommendation 2: Validation of the surrogate/final outcome relationship is needed. Recommendation 3: When using surrogate endpoints in cost-effectiveness modelling make it transparent, explore uncertainty and give advice on future research needed on the surrogate/final endpoint relationship.
Bucher HC. Studien mit Surrogatendpunkten: Nutzen und grenzen in der klinishen Entscheidungsfindung. Internist 2008; 49(6): 681-7.	Review paper	Ideally clinicians should base their treatment decisions on results from randomised controlled trials which include patient-important outcomes, such as quality of life, prevented disease events or death. However, in many countries drugs are approved based on data from surrogate endpoint trials. Recently, a controversy evolved on the reliability of results generated from these trials driven by unanticipated	Recent examples and different criteria on how clinicians can critically evaluate the validity of claims by experts or the pharmaceutical industry in regard to the expected patients' benefit from drugs approved by results from surrogate endpoint trials are presented.

		side effects or severe toxicity leading to the withdrawal of drugs that were solely approved based on evidence from surrogate endpoint trials.	
Helfand M, Balshem H. AHRQ Series Paper 2: Principles for developing guidance: AHRQ and the Effective Health Care Program. J Clin Epidemiol 2010; 63, 484-490.	Guidance document from AHRQ	Need for an analytic framework that makes it clear what surrogate outcomes may represent the final health outcomes and what bodies of evidence that link the surrogate outcomes to health outcomes.	Need for a strong link between surrogate outcome and final health outcome
Allgemeine Methoden Entwurf für Version 3.0 vom 15.11.2007 : Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG)	Methodological guidline for HTA or coverage decisions	There has to be a plausible, strong, consistent and in the same direction change both in the surrogate endpoint and in the clinical endpoint before surrogate endpoints can be accepted for coverage decisions.	Validation of the surrogate/final outcome relationship is needed.
Mangiapane S & Velasco Garrido M,Use of surrogate endpoints in HTA. GMS health technology assessment 2009, Vol. 5.	Systematic Review	*The study aimed to answer the following questions:* *Which criteria need to be fulfilled for a surrogate parameter to be considered a valid endpoint?* Which methods have been described in the literature for the assessment of the validity of surrogate endpoints? Which methodological recommendations concerning the use of surrogate endpoints have been made by international HTA agencies? Which placehas been given to surrogate endpoints in international and German HTA reports?	The analysis of methodological guidelines shows a very cautious position of HTA institutions regarding the use of surrogate endpoints in technology assessment. Surrogate endpoints have not been prominently used in HTA reports. None of the analysed reports based its conclusions solely on the results of surrogate endpoints. The analysis of German HTA reports show a similar pattern. The validation of a surrogate endpoint requires extensive research, including randomized controlled trials (RCT) assessing clinical relevant endpoints. The validity of a surrogate parameter is rather technology specific than disease-specific. Thus – even in the case of apparently similar technologies – it is necessary to validate the surrogate for every single technology (i. e. for

			every single active agent). The use of surrogate endpoints in the assessment of the benefit of health technologies is still to be seen very critically.
Velasco Garrido M & Mangiapane S, Surrogate outcomes in health technology assessment: An international comparison. IJTAHC, 25:3 (2009), 315-22.	Systematic Review	The study aimed to review the recommendations given by health technology assessment (HTA) institutions in their methodological guidelines concerning the use of surrogate outcomes in their assessments. In a second step, the study aimed at quantifying the role surrogate parameters take in assessment reports.	The authors identified thirty-four methodological guidelines, twenty of them addressing the issue of outcome parameter choice and the problematic of surrogate outcomes. Overall HTA agencies call on caution regarding the reliance on surrogate outcomes. None of the agencies has provided a list or catalog of acceptable and validated surrogate outcomes. Surrogate endpoints were used in 62 percent of the reports. However, only 3.6 percent were based upon surrogate outcomes exclusively. All of them assessed diagnostic or screening technologies and the surrogate outcomes were predominantly test characteristics.
Report on the Surrogate to Final Outcome Working Group to the Pharmaceutical Benefits Advisory Committee	Guidance document	The Working Group's aim was to develop a Framework for assessing the uncertainties associated with using an observed comparative treatment effect based on a surrogate to predict quantitatively as well as qualitatively a comparative treatment effect on a more patient-relevant outcome, particularly for incorporation into an economic evaluation.	The core of the Framework defines the information needed to consider the specific issues involved in assessing a proposed surrogate measure and its use to predict the effectiveness of a new medicine on its target clinical outcome particularly for incorporation into an economic evaluation.
Articles assessing the use of surrogate endpoints for marketing authorisation			
NOTE FOR GUIDANCE ON GENERAL CONSIDERATIONS FOR CLINICAL TRIALS (CPMP/ICH/291/95)	Guidance document	Guidance on response variables/endpoints in clinical trials from EMA	Surrogate endpoints may be used as primary endpoints when appropriate (when the surrogate is reasonably likely or well known to predict

			clinical outcome).
Report on the EMEA/CHMP Biomarkers Workshop. Doc Ref: EMEA/522496/2006	Worksho report	A workshop was held between EFPIA and CHMP/EMA	More fundamental scientific and clinical knowledge is required for the qualification of biomarkers as surrogate endpoints.
Final report from the EMEA/CHMP-think-tank group on innovative drug development. EMEA/127318/2007	Meeting report	The question on how to qualify biomarkers and surrogate endpoints was one of the most important scientific topics in the meeting	Before a biomarker can be accepted as a surrogate endpoint, there is a need to have confidence that changes in the marker reliably predict the desired clinical endpoint.
Fleming TR. Surrogate endpoints and FDAs accelerated approval process. Health Affairs, 24, no.1, 2005:67-78.	Discussi on paper	This paper considers issues related to validating surrogate endpoints—that is, identifying when effects on biological markers would accurately predict when treatment truly provides tangible benefit to patients. It proposes an endpoint hierarchy characterizing the relative reliability of outcome measures when used to evaluate clinical benefit. Finally, it considers the controversial issues in the implementation of the Food and Drug Administration's (FDA's) accelerated-approval process, where treatments only known to be biologically active can be marketed to the public while scientific trials are under way to determine whether these agents truly are more effective than toxic.	The paper suggests an endpoint hierarchy for outcome measures: Level 1: a true clinical-efficacy measure; Level 2: a validated surrogate endpoint (for a specific disease setting and class of interventions); Level 3: a nonvalidated surrogate endpoint, yet one established to be reasonably likely to predict clinical benefit (for a specific disease setting and class of interventions); Level 4: a correlate that is a measure of biological activity but that has not been established to be at a higher level The paper suggest to use only outcome measures that are level 1 or level 2 endpoints in phase 3 clinical trials.

Thomas R. Fleming, PhD, and David L. DeMets, PhD. Surrogate End Points in Clinical Trials: Are We Being Misled? Ann Intern Med. 1996;125:605-613.	Review article	An article aimed at discussing the use of surrogate endpoints in clinical trials.	Surrogate endpoints can be useful in phase 2 screening trials for identifying whether a new intervention is biologically active and for guiding decisions about whether the intervention is promising enough to justify a large definitive trial with clinically meaningful outcomes. In definitive phase 3 trials, except for rare circumstances in which the validity of the surrogate endpoint has already been rigorously established, the primary endpoint should be the true clinical outcome.
IOM (Institute of Medicine). 2011. Perspectives on biomarker and surrogate endpoint evaluation: Discussion forum summary. Washington, DC: The National Academies Press.	Policy paper	The aim of this discussion paper is to describe the framework for effective biomarker evaluation for use in marketing authorisation processes.	The biomarker evaluation process should consist of the following three steps: 1a. Analytical validation: analyses of available evidence on the analytical performance of an assay; 1b. Qualification: assessment of available evidence on associations between the biomarker and disease states, including data showing effects of interventions on both the biomarker and clinical outcomes; and 1c. Utilization: contextual analysis based on the specific use proposed and the applicability of available evidence to this use. This includes a determination of whether the analytical validation and qualification conducted provide sufficient support for the use proposed.