

IQWiG Reports – Commission No. A10-05

# **Validity of surrogate endpoints in oncology<sup>1</sup>**

## **Executive Summary**

---

<sup>1</sup> Translation of the executive summary of the rapid report A10-05 “Aussagekraft von Surrogatendpunkten in der Onkologie” (Version 1.1; Status: 21.11.2011). Please note: This translation is provided as a service by IQWiG to English-language readers. However, solely the German original text is absolutely authoritative and legally binding.

# Publishing details

**Publisher:**

Institute for Quality and Efficiency in Health Care

**Topic:**

Validity of surrogate endpoints in oncology

**Contracting agency:**

Federal Joint Committee

**Commission awarded on:**

19.08.2010

**Internal Commission No.:**

A10-05

**Address of publisher:**

Institute for Quality and Efficiency in Health Care  
Dillenburger Str. 27  
51105 Cologne  
Germany

Tel: +49-(0)221/35685-0

Fax: +49-(0)221/35685-1

E-mail: [berichte@iqwig.de](mailto:berichte@iqwig.de)

Website: [www.iqwig.de](http://www.iqwig.de)

## **Background**

With its letter of 19.08.2010 the Federal Joint Committee (G-BA) commissioned the Institute for Quality and Efficiency in Health Care (IQWiG) to assess the validity of surrogate endpoints in oncology.

## **Research question**

The primary goal of this research was to present recommendations for the validation of surrogate endpoints in oncology.

These recommendations were to be inferred from the results of the following 2 sub-goals:

- 1) Summarizing presentation and assessment of methodological approaches aiming to validate surrogate endpoints
- 2) Presentation and assessment of studies on the validation of tumour response parameters as surrogates for patient-relevant endpoints, in particular for patient survival. (The work on sub-goal 2 was conducted using colon and breast cancer as examples.)

## **Methods**

A literature search was performed for each of the two sub-goals to identify relevant publications.

The search for relevant publications on sub-goal 1 (presentation and assessment of methodological approaches) was conducted in MEDLINE, EMBASE, and the Cochrane Methodology Register (Methods Studies). The period from 2007 to September 2010 was covered; the period before 2007 was covered by other systematic reviews that were identified in the preliminary search and considered in the assessment. The publications included were classified according to methodological aspects and described in summary. The requirements for surrogate validation were inferred from these findings and presented in summary.

The search for relevant publications on sub-goal 2 (assessment of validation studies for surrogate endpoints used in colon and breast cancer) was also performed in MEDLINE, EMBASE, and the Cochrane Methodology Register (Methods Studies). The period up to August 2010 (colon cancer) and October 2010 (breast cancer) was covered. In a first step the reliability of the validation studies was assessed. In this context, the term “reliability” comprises the following criteria:

- Application of a recognized approach described in the scientific literature
- Conduct of analyses to test the robustness and generalizability of results
- Systematic compilation of the underlying data for the validation
- Sufficient restriction of indications or degrees of disease severity
- Sufficient restriction of the interventions investigated

- Clear definitions of the endpoints investigated

On the basis of these criteria, the reliability of the validation studies was in each case classified as high, limited, moderate or low.

Subsequently the results for the validity of the investigated surrogate endpoints were analysed.

## Results

### *Sub-goal 1: Presentation and assessment of methodological approaches*

The systematic search for publications on surrogate validation showed that suitable proposals on methodological approaches for the validation of surrogate endpoints are available, which have also been tested in practice. Therefore, in principle such a validation seems possible. The following basic conclusions can be inferred from the available literature:

- 1) The demonstration of a correlation between the surrogate and the clinical endpoint is in itself insufficient to validate a surrogate.
- 2) It is not readily possible to transfer conclusions about the validity of surrogates between different diseases or disease grades or between different interventions. The validity of a surrogate can be both disease-specific and intervention-specific, i.e. a conclusive validation of a surrogate by means of statistical methods can possibly only be performed within an indication or intervention. It must be tested and justified to what extent conclusions about the validity of surrogates can be transferred between different diseases or interventions. In order to assess transferability in validation studies that include various disease entities or interventions, at least analyses on heterogeneity should be available.
- 3) Comprehensive data are required to validate a surrogate, preferably a meta-analysis of several randomized trials showing sufficient certainty of results. Both the surrogate and the clinical endpoint must have been analysed in these studies.
- 4) There is neither a universally applicable measure for surrogate validation nor a general best estimation method nor a generally accepted threshold which, if exceeded, would mean proof of validity.
- 5) In current practice, correlation-based approaches are primarily applied for surrogate validation, with estimation of correlation measures on a study level and an individual level. In order to demonstrate validity, on a study level the pertinent literature usually calls for a high correlation between effects on the surrogate and the patient-relevant endpoint; various authors propose a correlation of 0.9 as a potential threshold.

Particularly in cases where (under consideration of the corresponding confidence intervals) no high correlation is evident,<sup>2</sup> i.e. the validity of the surrogate remains unclear, conclusions

---

<sup>2</sup> lower confidence limit of the estimated correlation coefficient < 0.85

about patient-relevant endpoints can still be made by applying the surrogate threshold effect (STE) concept according to Burzykowski & Buyse (2005, 2006) if sufficiently large effects on the surrogate have been shown. The STE concept allows to specify threshold values for the decision as to whether an observed effect on the surrogate is also accompanied (with sufficient certainty) by an effect on the endpoint of interest. To draw such a conclusion, the lower confidence limit of the treatment effect on the surrogate must be larger than the STE.

The application of non-correlation-based approaches needs sufficient justification, in particular if only one study is to be used as the data basis.

*Sub-goal 2: Assessment of validation studies for surrogate endpoints used in colon and breast cancer*

The systematic search for studies on surrogate validation in colon and breast cancer yielded a total of 15 and 6 relevant publications on colon and breast cancer, respectively.

In summary it can be concluded that the included validation studies by and large confirm the results of the methods review. However, none of the studies consistently fulfilled recognized quality criteria. This may also be due to the fact that some of the validation studies were not primarily targeted towards surrogate validation, but represented methodological papers on the application of validation methods. Accordingly, various papers from sub-goal 2 were also included in the methods review (sub-goal 1). In the validation studies different interventions or indications were often analysed conjointly, or it remained unclear whether the data basis was restricted to one intervention or indication. The transferability of the results of the surrogate validation between the included interventions or indications was not addressed. It also mostly remained unclear how the data basis had been compiled or its compilation proved to be non-systematic. Overall the reliability of the validation studies was low.

The validity of the investigated surrogates mostly remained unclear, also due to the statistical results of the validation studies. Disease-free survival (DFS) in early-stage colon cancer was an exception; here, in both validation studies a high correlation was shown. However, the reliability of these two studies was also low, so that no proof of the validity of DFS for overall survival can be inferred from these studies.

In summary, in the present assessment the validity of tumour response parameters as surrogates for patient-relevant endpoints (in this case overall survival) in colon and breast cancer remains unclear.

**Proposal for handling surrogate endpoints in benefit assessments of oncology therapies**

With the present report, IQWiG presents for discussion an algorithm for handling surrogate endpoints in benefit assessments of oncology therapies. This algorithm was developed based on a systematic search for surrogate validation methods (sub-goal 1), as well as on a systematic assessment of validation studies for two example indications, i.e. colon and breast cancer (sub-goal 2). On the basis of the available data and information, this proposal aims to

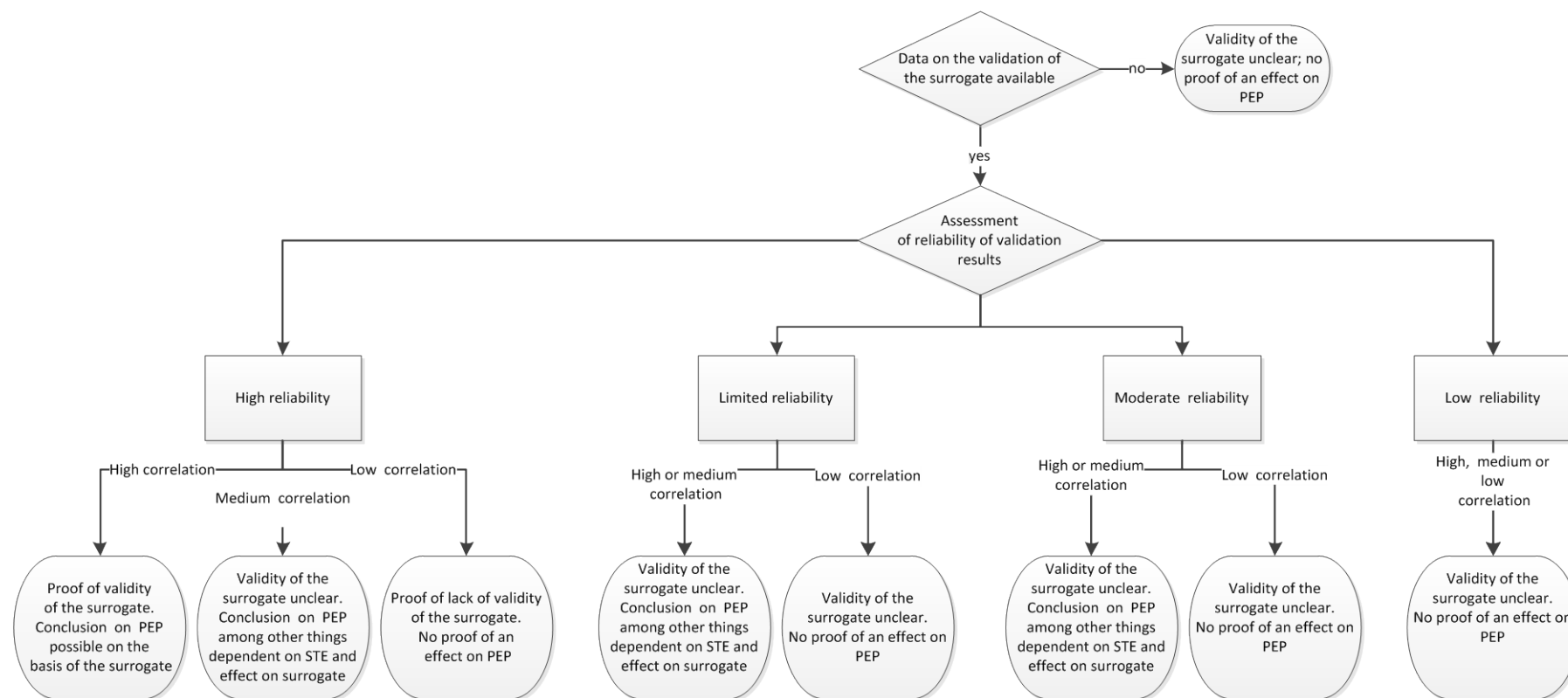
take a differentiated view of the findings for surrogate endpoints, i.e. a view that considers the certainty of results of the different constellations.

The results of the project are to serve the Federal Joint Committee as a basis for discussion regarding the utilization of surrogate endpoints in benefit assessments of oncological therapies.

Two levels need to be distinguished in the consideration of surrogate endpoints in benefit assessments. On the first level, initially the validity of the surrogate needs to be assessed. On the second level, the question arises as to what conclusions about patient-relevant endpoints can be drawn from the effects on the surrogate, depending on the categorization of the validity of the surrogate.

***Level 1: Assessment of the validity of the surrogate***

Figure A summarizes the deliberations on the approach to surrogate validation and on the connection of the validation results to the benefit assessment.



PEP: patient-relevant endpoint; STE: surrogate threshold effect

Figure A: Impact (of the reliability of validation studies and the correlation of the effects on the surrogate and patient-relevant endpoint) on the validity of the surrogate

In the case of high reliability of the validation study, proof of the validity of the surrogate can be inferred from a high correlation (lower limit of the 95% confidence interval)  $R \geq 0.85$  of the effects on the surrogate and the patient-relevant endpoint. Low correlation (upper limit of the 95% confidence interval)  $R \leq 0.7$  results in proof of the lack of validity of the surrogate. If correlation is medium ( $R < 0.85$  to  $> 0.7$ ) the validity of the surrogate remains unclear. The validity of the surrogate is also unclear, i.e. cannot be conclusively assessed, in the case of limited or moderate reliability of the validation study. If no data on surrogate validation are available, or if the data available only show low reliability, no conclusion about the patient-relevant endpoint can be drawn on the basis of the results for the surrogate. Thus no proof of an effect on the patient-relevant endpoint can be inferred.

The actual goal of surrogate validation is to answer the question as to what conclusions about patient-relevant endpoints can be drawn from surrogate endpoints (level 2). A direct inference can be made from effects measured for a valid surrogate to effects on the patient-relevant endpoint. Effects on surrogates that are proved invalid allow no conclusions about patient-relevant endpoints. If the validity of the surrogate is unclear, conclusions about patient-relevant endpoints can be made using the STE concept; these conclusions have a graded certainty of results and thus take the evidence base into account in a differentiated way. Below we present a proposal for the interpretation of effects on surrogates in cases of unclear validity.

***Level 2: Conclusions about patient-relevant endpoints drawn from effects on surrogates, depending on the categorization of the validity of the surrogate***

Figure B shows the first steps of the algorithm for the consideration of results for surrogate endpoints in benefit assessments.



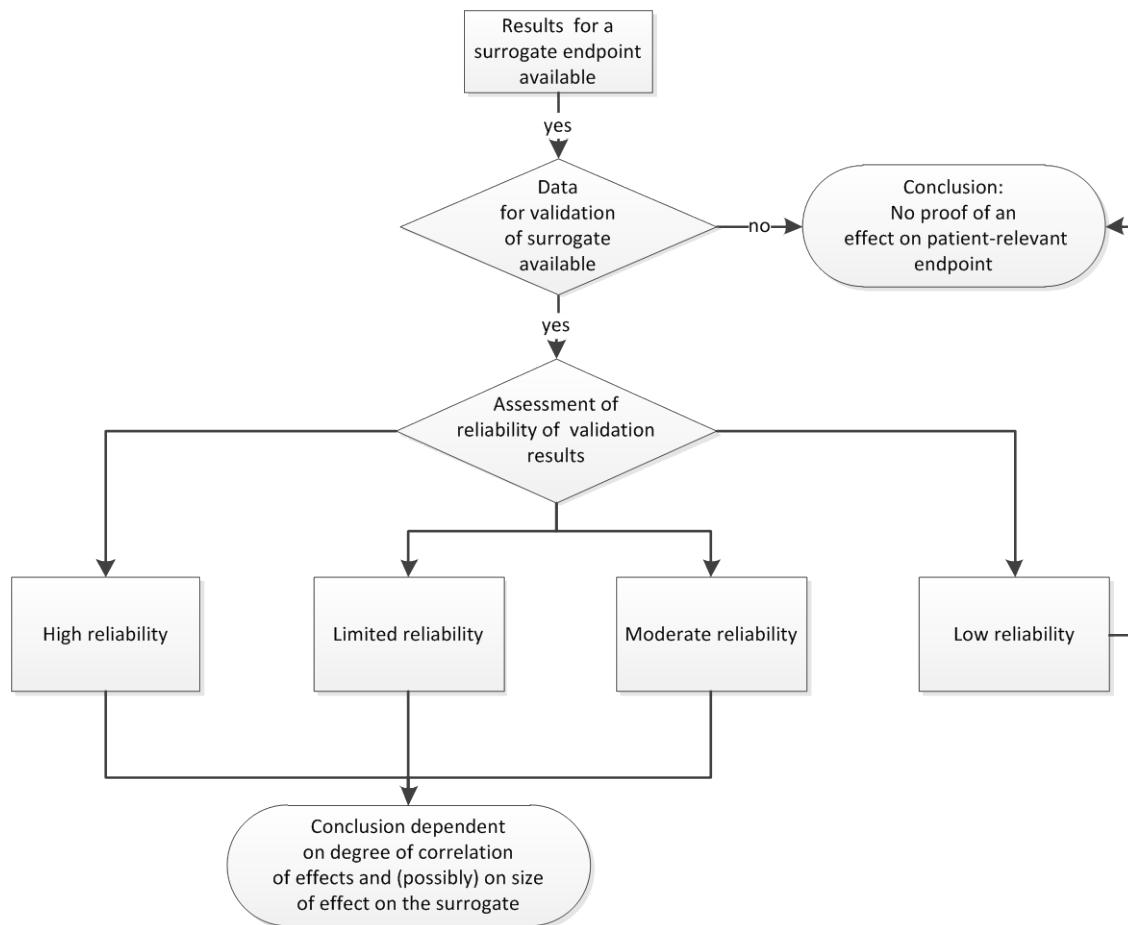
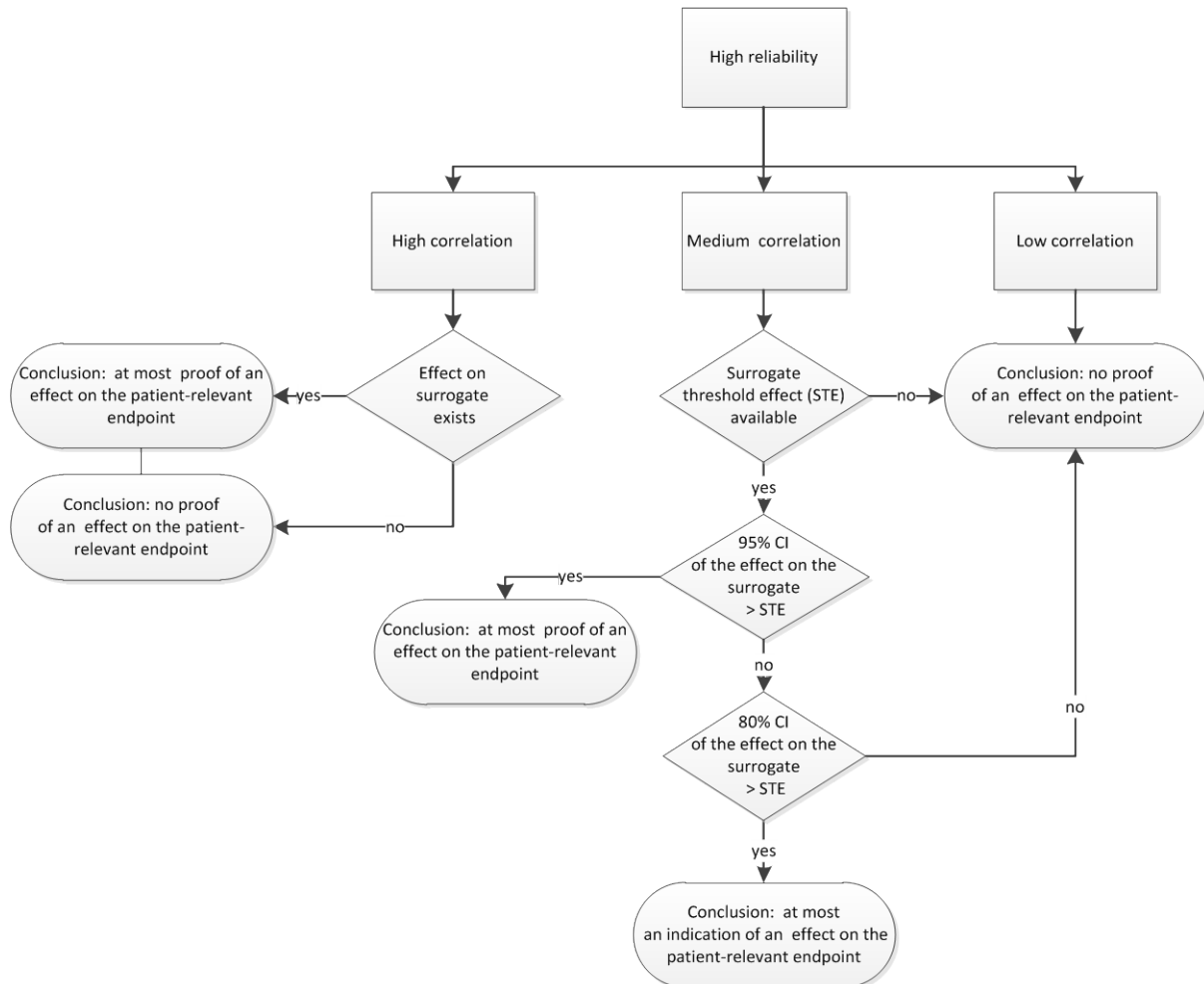


Figure B: Consideration of surrogate endpoints in benefit assessments: importance of validation studies and their reliability

If results for a surrogate endpoint are presented within the framework of a benefit assessment, the question arises as to whether data are available for surrogate validation. If this is not the case, the results for the surrogate endpoint do not provide proof (or an indication or hint) of an effect on a patient-relevant endpoint. If data on surrogate validation are available, in a first step they are assessed with regard to their reliability. In the case of low reliability of the validation results, categorically no proof of an effect on the patient-relevant endpoint can be inferred from the results for the surrogate endpoint. In the case of high, limited or moderate reliability of the validation results, the conclusions about patient-relevant endpoints are dependent on the degree of correlation of the effects on the surrogate and the patient-relevant endpoint in the validation studies. They are also (possibly) dependent on the size of the effect on the surrogate in the studies included in the benefit assessment (shortened to “surrogate(s) in the benefit assessment” in the following text). In this context, the basic principle is that the certainty of conclusions in the benefit assessment (e.g. proof or indication of an effect on a patient-relevant endpoint) is all the higher, the higher the reliability of the validation studies (and thus of the surrogate), (see Figures C to E).

*Conclusions for the benefit assessment in the case of high reliability of the validation studies*

Figure C shows the potential conclusions for the benefit assessment in the case of high reliability of the validation studies.



CI: confidence interval; STE: surrogate threshold effect

Figure C: Consideration of surrogate endpoints in benefit assessments: potential conclusions in the case of high reliability of the validation studies

Depending on the correlation between surrogate and patient-relevant endpoint in the validation studies, different conclusions are inferable from results for the surrogate endpoints in the benefit assessment.

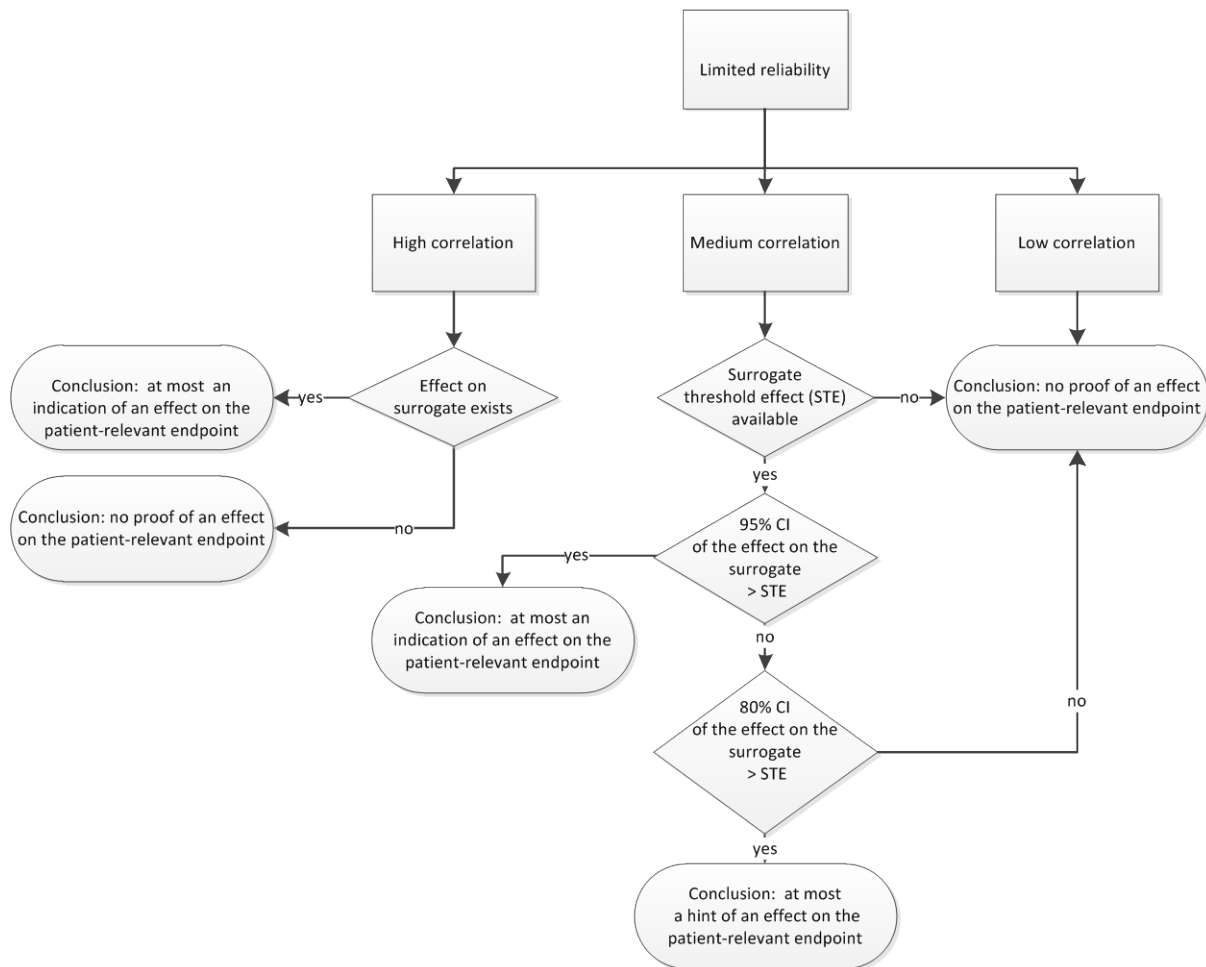
- From validation studies with a high reliability, in the case of high correlation and an effect on the surrogate it can at most be inferred that there is proof of an effect on the patient-relevant endpoint. This conclusion means that in addition to assessing the validity of the surrogate in validation studies, the certainty of results of data on the surrogate in the benefit assessment is considered. Depending on this certainty of results, the conclusion about the patient-relevant endpoint may be downgraded from “proof” to “an indication”.

If the studies included in the benefit assessment showed no effect of the intervention on the surrogate, i.e. if no statistically significant group difference was found in these studies or in a meta-analysis of these studies, in general an effect on the patient-relevant endpoint is not proven (this also applies to the path with medium correlation).

- In the case of high reliability of the validation studies and medium correlation between the surrogate and the patient-relevant endpoint in these studies, the effect on the surrogate in the benefit assessment is compared with the STE. If such an STE is not available, an effect on the patient-relevant endpoint is not proven. If an STE is available, initially the 95% CI of the effect on the surrogate in the benefit assessment is related to this STE. If the 95% CI lies fully above the STE, there is at most proof of an effect on the patient-relevant endpoint. (In the case of increased uncertainty of results in the studies included in the benefit assessment, e.g. due to a high risk of bias, this result can again be down-graded to an indication.) If the 95% CI does not lie fully above the STE, the 80% CI of the effect on the surrogate is related to the STE. If the 80% CI lies fully above the STE, there is at most an indication of an effect on the patient-relevant endpoint. If the 80% CI of the effect on the surrogate does not fully exceed the STE, in the case of high reliability of the validation studies and medium correlation, an effect on the patient-relevant endpoint is not proven, and neither is an indication or hint of such an effect available.
- In the case of high reliability of the validation studies and low correlation between the surrogate and patient-relevant endpoint, the surrogate is deemed to be not valid. Categorically no proof (nor indication nor hint) of an effect on the patient-relevant endpoint can be inferred from an effect on the surrogate.

*Conclusions for the benefit assessment in the case of limited reliability of the validation studies*

Figure D shows the potential conclusions for the benefit assessment in the case of limited reliability of the validation studies.



CI: confidence interval; STE: surrogate threshold effect

Figure D: Consideration of surrogate endpoints in benefit assessments: potential conclusions in the case of limited reliability of the validation studies

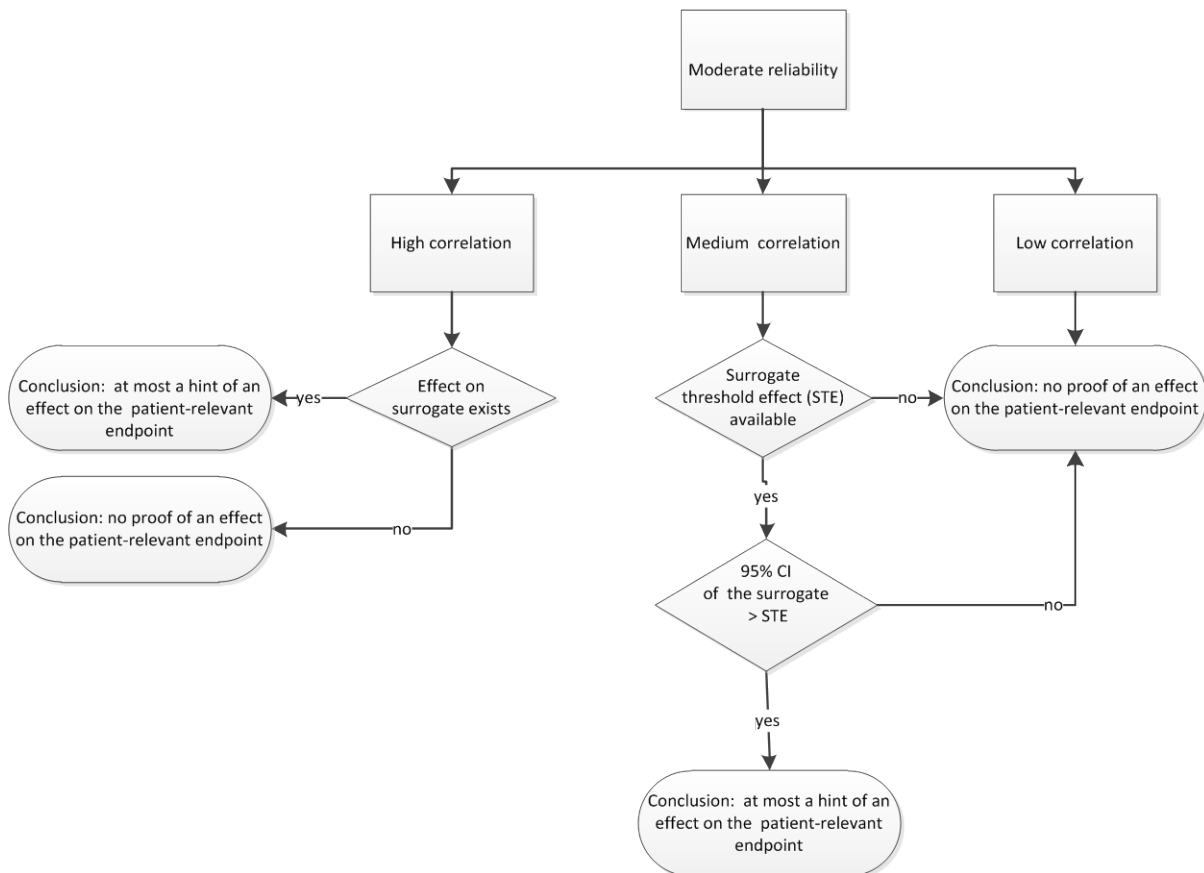
Compared to the potential conclusions in the benefit assessment in the case of high reliability of the validation studies, in cases of limited reliability of these studies, all conclusions about effects on patient-relevant endpoints have a lower certainty of results.

- In the case of limited reliability and high correlation of the surrogate and patient-relevant endpoint, it can at most be inferred from effects on surrogates in the benefit assessment that there is an indication of effects on patient-relevant endpoints.
- In the case of limited reliability and medium correlation, if the STE is exceeded by the 95% CI of the surrogate, an indication of an effect on the patient-relevant endpoint can at most be inferred. If the STE is exceeded by the 80% CI of the surrogate, a hint of an effect on the patient-relevant endpoint can at most be inferred.
- In the case of limited reliability and low correlation, categorically no proof is available of an effect on the patient-relevant endpoint.

For all final conclusions, additionally the certainty of results of studies included in the benefit assessment is taken into account as described above, i.e. a limitation in the certainty of results leads to a downgrading of the conclusion.

*Conclusions for the benefit assessment in the case of moderate reliability of the validation studies*

Figure E shows the potential conclusions for the benefit assessment in the case of moderate reliability of the validation studies.



CI: confidence interval; STE: surrogate threshold effect

Figure E: Consideration of surrogate endpoints in benefit assessments: potential conclusions in the case of moderate reliability of the validation studies

In the case of moderate reliability of the validation studies, results for surrogates in a benefit assessment can at most provide hints of effects on patient-relevant endpoints. In the case of medium correlation, for the derivation of such a conclusion, exclusively the 95% CI is related to the STE. The further procedure follows the approach used in the case of high or limited reliability of validation studies: after assessment of the conclusion that can at most be inferred concerning an effect on a patient-relevant endpoint, the certainty of results for the surrogate in

the benefit assessment is taken into account. In the case of a limitation of the certainty of results, the classification “hint” is downgraded to “no proof” (as well as “no indication” and “no hint”).

## **Conclusions**

### ***Methods for surrogate validation***

Suitable proposals on methodological approaches for the validation of surrogate endpoints are available, which have also been tested in practice. In principle a validation of surrogates therefore seems possible. In this context, there is a general international consensus with regard to the following basic conclusions:

- The demonstration of a correlation between the surrogate and the clinical endpoint is in itself insufficient to validate a surrogate.
- It is not readily possible to transfer conclusions about the validity of surrogates between different diseases or disease grades or between different interventions. In order to assess transferability in validation studies that include various disease entities or interventions, at least analyses on heterogeneity should be available.
- Comprehensive data are required to assess the validity of a surrogate, preferably a meta-analysis of several randomized trials showing sufficient certainty of results. Both the surrogate and the clinical endpoint must have been analysed in these studies.
- There is neither a universally applicable measure for surrogate validation, nor a general best estimation method, nor a generally accepted threshold which, if exceeded, would mean proof of validity.
- In current practice, correlation-based approaches are primarily applied for surrogate validation, with estimation of correlation measures on a study level and individual level. In this context the STE concept is helpful in determining threshold values for the decision process.

### ***Validation studies for colon and breast cancer***

The available validation studies on breast and colon cancer allow no final conclusion about the validity of tumour response parameters for overall survival in patients with these diseases. In most cases the correlation of the respective effects was not high enough, so that the prerequisite for proof of validity was not fulfilled. In addition, in all studies reliability was classified as low. Therefore, independent of the statistical results, the validity of all investigated surrogate endpoints ultimately remained unclear. Particularly the following factors were decisive for the low reliability of the validation studies:

- Surrogate validation was largely not intervention-specific and the impact of the different interventions on the overall result was not analysed.

- Different indications or different grades of disease severity were often summarized and the impact of the different treatment situations on the overall result was not analysed.
- Regarding colon cancer, the compilation of the data basis was often unsystematic or the approach unclear.

Overall, the validation studies show that in principle these deficiencies are amendable. It would be in the interests of the most informed treatment possible for patients with cancer if validation studies were conducted systematically and reported transparently, taking into account the methodological standards mentioned. This requires the combined efforts of all parties involved. The associated improvement in knowledge on the relevance of surrogate endpoints would also be suited to show potential decision paths for the assessment of new oncology drugs, despite the limited evidence on these drugs.

### ***IQWiG's proposal on the handling of surrogate endpoints in benefit assessments of oncological therapies***

Two levels need to be distinguished in the consideration of surrogate endpoints in benefit assessments. On the first level, initially the validity of the surrogate needs to be assessed. On the second level the question arises as to what conclusions about patient-relevant endpoints can be drawn from the effects on the surrogate, depending on the categorization of the validity of the surrogate.

If results for a surrogate endpoint are presented within the framework of a benefit assessment, it should first be assessed whether data for surrogate validation are available. If this is not the case, the results for the surrogate endpoint do not provide proof (nor an indication nor hint) of an effect on the patient-relevant endpoint. If data on surrogate validation are available, in a first step they are assessed with regard to their reliability. In the case of low reliability of the validation results, categorically no proof of an effect on the patient-relevant endpoint can be inferred from the results for the surrogate endpoint. In the case of high, limited or moderate reliability of the validation results, the conclusions about patient-relevant endpoints are dependent on the degree of correlation of the effects on the surrogate and the patient-relevant endpoint in the validation studies. They are also (possibly) dependent on the size of the effect on the surrogate in the benefit assessment. For the derivation of such conclusions, the effect on the surrogate shown in the studies included in the benefit assessment is compared to an STE. In this context, the basic principle is that the certainty of conclusions in the benefit assessment (e.g. proof or indication of an effect on a patient-relevant endpoint) is all the higher, the more certainly an effect on the surrogate also represents an effect on the patient-relevant endpoint.

**Keywords:** biological markers; data interpretation, statistical; predictive value of tests; medical oncology; breast neoplasms; colorectal neoplasms; systematic review

*The full report (German version) is published under [www.iqwig.de](http://www.iqwig.de)*