



AKADÉMIAI KIADÓ

Binary surrogate endpoints in clinical trials from the perspective of case definitions

ANDREAS HAHN¹, ANDREAS PODBIELSKI¹,
MARKUS M. HEIMESAAT², HAGEN FRICKMANN^{1,3†} and
PHILIPP WARNKE^{1†*}

European Journal of
Microbiology and
Immunology

11 (2021) 1, 18-22

DOI:
[10.1556/1886.2020.00031](https://doi.org/10.1556/1886.2020.00031)
© 2020 The Author(s)

¹ Institute for Medical Microbiology, Virology and Hygiene, University Medicine Rostock, Rostock, Germany

² Institute of Microbiology, Infectious Diseases and Immunology, Charité – University Medicine Berlin, Berlin, Germany

³ Department of Microbiology and Hospital Hygiene, Bundeswehr Hospital Hamburg, 20359 Hamburg, Germany

Received: November 22, 2020 • Accepted: December 7, 2020
Published online: March 4, 2021

ORIGINAL RESEARCH PAPER



ABSTRACT

Introduction: Surrogate endpoints are widely used in clinical trials, especially in situations where the endpoint of interest is not directly observable or to avoid long trial periods. A typical example for this case is frequently found in clinical trials in oncology, where overall survival (OS) as endpoint of interest and progression free survival (PFS) as surrogate endpoint are discriminated. **Methods:** Based on the perspective of case definitions on surrogate endpoints, we provide a formal definition of such endpoints followed by a description of the structure of surrogate endpoints. **Results:** Surrogate endpoints can be considered as case definitions for the endpoint of interest. Therefore, the performance of surrogate endpoints can be described using the classical terminology of diagnostic tests including sensitivity and specificity. Since such endpoints always focus on sensitivity with necessarily reduced specificity, efficacy estimates based on such endpoints are in general biased. **Conclusion:** The abovementioned has to be taken into account while interpreting the results of clinical trials and should not be ignored while planning or conducting a study.

KEYWORDS

surrogate endpoints, bias, clinical trial, case definition

BACKGROUND

Surrogate endpoints are common substitutes for true endpoints in clinical trials and frequently used to shorten the observation time [1–3]. Nevertheless, they have been controversially discussed, especially because clear standards for their validation are still missing [4–6].

But prior to a validation, there should be a clear and formal definition of a surrogate endpoint. So far, there are some more or less clear explanations of a surrogate endpoint, e.g., “The purpose of a surrogate endpoint is to draw conclusions about the effect of intervention on true endpoint without having to observe the true endpoint.” [7] or “A surrogate endpoint of a clinical trial is a laboratory measurement or a physical sign as a substitute for a clinical meaningful endpoint that measures directly how a patient feels, functions or survives. Changes induced by a therapy on a surrogate endpoint are expected to reflect changes in a clinical meaningful endpoint” [8].

Common to all definitions is the requirement that the surrogate endpoint is predictive for the true endpoint. In addition, there is the expectation that a surrogate endpoint shall reduce the observation time by increasing endpoint sensitivity. Oncological trials are standard

†Hagen Frickmann and Philipp Warnke contributed equally.

*Corresponding author.
E-mail: philipp.warnke@med.uni-rostock.de

examples for this intention. Thereby, progression free survival (PFS) is used as a surrogate endpoint for overall survival (OS). PFS is defined as the time from randomization to disease progression or death of any cause (whatever occurs first) while OS is simply defined as time from randomization to death of any cause. Obviously, PFS is very sensitive regarding OS and therefore reduces the observation time. But a standard approach for a general validation of the main goal, i.e., predictiveness for OS, is still missing.

From the perspective of case definitions, we provide a formal definition of surrogate endpoints regarding the two goals predictivity and shortening of the observation time. Then, we apply published adjusted estimators for prevalence and intervention efficacy to identify a category of surrogate endpoints that could be used in clinical trials without validation. Subsequently, we use the very common example of PFS and OS to discuss our conclusions on the structure of surrogate endpoints.

METHODS

The primary requirement of a surrogate endpoint is the assessment of an endpoint that is sufficient (predictive) for the true endpoint. This indicates that:

$$\text{Surrogate endpoint} \Rightarrow \text{True endpoint}$$

The main intention for the establishment of surrogate endpoints is an increase of cases which makes it easier to reach a study endpoint. In this situation, reaching a surrogate endpoint is a necessary condition for reaching the true endpoint but not a sufficient one:

$$\text{Surrogate endpoint} \Leftarrow \text{True endpoint}$$

For each true endpoint, its surrogate endpoint can be understood as a case definition of the true endpoint. A formal definition of a case definition was introduced by Hahn et al. [9] and is given by:

Let Ω be the set of all symptoms and attributes and $P(\Omega)$ its power set. Then, every subset $C \subseteq P(\Omega)$ is called a case definition C .

For every case definition (e.g., for a disease) and a given set of individuals, a case is an individual whose symptoms and attributes fulfill the case definition:

Let I be the set of all individuals. $S_k \in P(\Omega)$ denotes the individual set of symptoms and attributes for individual $i_k \in I$. A case is defined by Eq. (1):

$$\text{with } \begin{cases} \text{Case} : I \times P(\Omega) \rightarrow \{0, 1\} \\ \text{Case}(i_k, S_k) = \begin{cases} 1 & \text{if } \exists c \in C : c \subseteq S_k \\ 0 & \text{else} \end{cases} \end{cases} \quad (1)$$

Conducting this definition, we can formally define three categories of surrogate endpoints:

Let C_S and C_T be two case definitions.

– C_S is called first category surrogate for C_T , if holds:

$$\{i_l : \text{Case}_S(i_l, S_S) = 1\} \subseteq \{i_l : \text{Case}_T(i_l, S_T) = 1\} \\ = P(\text{Case}_T = 1 | \text{Case}_S = 1) = 1 \quad (2)$$

– C_S is called second category surrogate for C_T , if holds:

$$\{i_k : \text{Case}_T(i_k, S_T) = 1\} \subseteq \{i_l : \text{Case}_S(i_l, S_S) = 1\} \\ = P(\text{Case}_T = 0 | \text{Case}_S = 0) = 1$$

– C_S is called third category surrogate, if it is neither of the first nor of the second category.

First category surrogate endpoints are sufficient for the true endpoint. Second category surrogate endpoints are necessary but not sufficient. Third category surrogate endpoints are neither sufficient nor necessary for the true endpoints. A visualization of the three categories of surrogate endpoints is given by Fig. 1.

Obviously, a surrogate endpoint has diagnostic characteristics regarding the endpoint of interest with sensitivity Se and specificity Sp :

$$Se_S = P(\text{Case}_S = 1 | \text{Case}_T = 1) \quad (3)$$

$$Sp_S = P(\text{Case}_S = 0 | \text{Case}_T = 0)$$

Based on Eq. (2) follows that the specificity of first category surrogate endpoints is equal to one while the sensitivity is ≤ 1 . For second category surrogate endpoints holds that they have a sensitivity that is equal to one while the specificity is ≤ 1 . For third category surrogate endpoints holds that both sensitivity and specificity are < 1 .

For the rate of true cases that is assessed by a surrogate endpoint, an unbiased point estimator is given by [10, 11]:

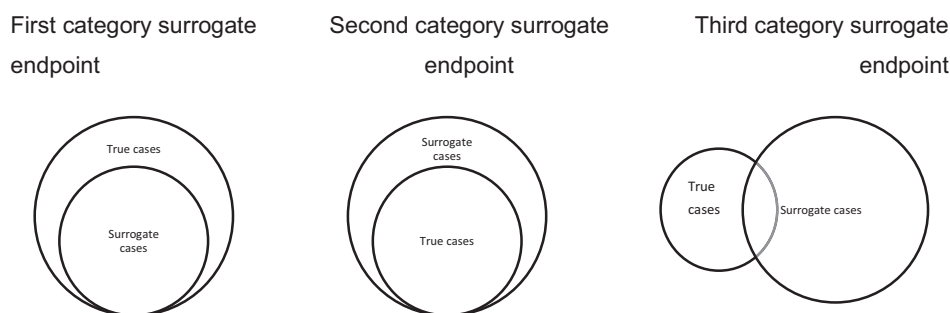


Fig. 1. Three categories of surrogate endpoint visualized by a Venn diagram

$$Rate_{True\ endpoint} = \frac{\text{rate of surrogate cases} - 1 + Sp}{Se + Sp - 1} \quad (4)$$

Based on Eq. (4), a consistent estimator for the rate ratio of true endpoints – if the diagnostic performance of surrogate endpoints does not vary over study arms – is given by [11]:

$$RR_{True\ endpoint} = \frac{\text{rate of surrogate cases}_1 - 1 + Sp}{\text{rate of surrogate cases}_2 - 1 + Sp} \quad (5)$$

It has also to be taken into account that surrogate endpoints are always time-related with regard to the endpoint of interest. This relation can be prospective, retrospective as well as contemporaneous.

RESULTS

Based on the definitions and equations above, the following results can be postulated for the application of surrogate endpoints in clinical and epidemiological trials.

With Eq. (4) follows that for the estimation of rates like prevalence rates, all categories of surrogate endpoints are biased. With Eq. (5) follows that only first category surrogate endpoints are consistent when applying unadjusted rates of the surrogate endpoint while the rates of second and third category surrogate endpoints have to be adjusted for their diagnostic performance to avoid this bias even for large sample sizes:

$$\begin{aligned} \text{First category surrogate endpoint: } RR_{True\ endpoint} &= \frac{\text{rate of surrogate cases}_1}{\text{rate of surrogate cases}_2} \\ \text{Second and third category surrogate endpoints: } RR_{True\ endpoint} &= \frac{\text{rate of surrogate cases}_1 - 1 + Sp}{\text{rate of surrogate cases}_2 - 1 + Sp} \end{aligned} \quad (6)$$

It has to be noted that this only holds for the situation that the diagnostic performance of the surrogate endpoints is constant over study arms. This, of course, can in general only be assumed for double blinded randomized clinical trials. In all other cases holds:

$$\begin{aligned} \text{First category surrogate endpoint: } RR_{True\ endpoint} &= \frac{Se_2(\text{rate of surrogate cases}_1)}{Se_1(\text{rate of surrogate cases}_2)} \\ \text{Second and third category surrogate endpoints: } RR_{True\ endpoint} &= \frac{Sp_2(\text{rate of surrogate cases}_1 - 1 + Sp_1)}{Sp_1(\text{rate of surrogate cases}_2 - 1 + Sp_2)} \end{aligned} \quad (7)$$

One of the most popular examples is PFS as surrogate for OS in oncological trials, especially in palliative settings. In curative settings, Complete Response (CR) is often used as surrogate for curation, especially in the case of Sustainable Complete Response (SCR).

PFS is a composite endpoint of progression or death of any cause. Commonly, it is interpreted as predictive for death of any cause and therefore used as a surrogate for the overall survival. At the time point of evaluation, PFS is obviously a surrogate of the second category for OS, because every study subject reaching OS endpoints also reaches the PFS endpoint (because the PFS endpoint is a combination of progression and overall survival). But with this interpretation, PFS is worthless, because OS as the endpoint of interest is measured at the same time. So, PFS makes only sense if it is interpreted as a “look into the future” regarding overall survival. But with this interpretation, PFS only fulfills the definition of a third category endpoint, because progression is not a necessary condition for death of any cause but death after progression. This means that reaching the surrogate endpoint is neither necessary nor sufficient for reaching the endpoint of interest.

In curative settings, CR as well as SCR (Complete Response with a duration of at least a minimal time period or CR until the end of observation) are also often interpreted as surrogates for curation, because a follow-up time of 5 or more years is very rare at the time of marketing authorization. Obviously, there is no curation without CR but of course there can be cases of relapse after CR, if “curation” is defined as complete response until death. In this way, CR as well as SCR are both surrogate endpoints of the second category, because the reaching of this endpoint is necessary

for curation. Since the SCR is more specific than the very sensitive CR alone, it is superior to the less specific surrogate.

Summarized, only the rates of first category surrogate endpoints can be used without adjustment to get consistent estimations for the intervention efficacy regarding the true

endpoint. And even this only holds if the diagnostic performance of the surrogate endpoint does not vary over study arms. This, of course, can only be assumed for double blinded randomized clinical trials. The high specificity of first category surrogate endpoints is in contrast with the intention to establish surrogate endpoints, which is increasing the sensitivity of an endpoint to increase the number of cases.

DISCUSSION

Surrogate endpoints are very popular in clinical trials, especially if the true endpoint of interest cannot be observed within acceptable timelines [12]. In consequence, this leads to surrogate endpoints that inflate the number of cases in order to reduce the observation time. The key assumption of surrogate endpoints is, on the other hand, that they are predictive for the endpoint of interest [8]. For the first time from the perspective of case definitions, we have therefore proposed a formal definition of surrogate endpoints with three categories. The first category represents the key assumption and is in contrast to the second category, which represents the intention of shortening the observation time of clinical trials. The third category represents all surrogate endpoints that are neither of the first nor of the second category. Based on the application of a published consistent intervention efficacy estimator [13], we conclude that only first category surrogate endpoints can be used without further evaluation for their diagnostic performance characteristics regarding the endpoint of interest. Since surrogate endpoints of this category are specific case definitions of the true endpoint, they are in contrast to the intention of many oncological trials of reducing the observation time by increasing the number of cases. Instead of this, they lead to decreasing case numbers, making them unfeasible for this intention.

We further conclude that surrogate endpoints of the second and third category cannot be predictive by themselves because they are not sufficient for the endpoint of interest. Therefore, it is not surprising that the authors of a recent publication [14] conclude: “Using breast cancer as an example, we evaluated the underlying evidence for the surrogate endpoints for solid tumors listed in the FDA’s Table of Surrogate Endpoints and found weak or missing correlations of treatment effects on these surrogates with treatment effects on OS. Surrogate measures should be predictive of clinical benefit to be useful in supporting regular FDA approval.”

In addition, the results of our modeling only hold for trials where the diagnostic performance of the surrogate endpoint does not vary over study arms, which can only be assumed for double blinded randomized trials. If this cannot be assumed, even surrogate endpoints of the first category are likely to be biased regarding the endpoint of interest.

CONCLUSION

Overall, surrogate endpoints should be applied, defined and interpreted with caution and the use of surrogate endpoints with the intention to shorten the observation timelines by

inflating the number of cases should be avoided. If feasible with realistic effort, the true endpoints should be assessed.

Funding: No financial support was received for this study.

Authors’ contribution: AH and HF jointly planned the study. AH performed the modeling and wrote the manuscript. All authors have jointly optimized and reviewed the manuscript.

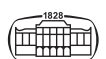
Conflict of interest: Nothing to declare.

LIST OF ABBREVIATIONS

CR	complete response
OS	overall survival
PFS	progression free survival
SCR	sustainable complete response

REFERENCES

1. American Society of Clinical Oncology. Outcomes of cancer treatment for technology assessment and cancer treatment guidelines. *J Clin Oncol* 1996;14:671–9.
2. Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? *Ann Intern Med* 1996;125:605–13.
3. Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. *Stat Med* 1989;8:431–40.
4. Ciani O, Buyse M, Garside R, Pavey T, Stein K, Sterne JA, et al. Comparison of treatment effect sizes associated with surrogate and final patient relevant outcomes in randomised controlled trials: meta-epidemiological study. *BMJ* 2013;346:f457.
5. Götzsche PC, Liberati A, Torri V, Rossetti L. Beware of surrogate outcome measures. *Int J Technol Assess Health Care* 1996;12:238–46.
6. Grimes DA, Schulz KF. Surrogate end points in clinical research: hazardous to your health. *Obstet Gynecol* 2005;105(5 Pt 1):1114–8.
7. Baker SG. Surrogate endpoints: wishful thinking or reality? *J Natl Cancer Inst* 2006;98:502–3.
8. Temple RJ. A regulatory authority’s opinion about surrogate endpoints. In: Nimmo WS, Tucker GT, editors. *Clinical measurement in drug evaluation*. New York: J. Wiley; 1995.
9. Hahn A, Frickmann H, Zautner AE. Impact of case definitions on efficacy estimation in clinical trials-A proof-of-principle based on historical examples. *Antibiotics (Basel)* 2020;9:7.
10. Gart JJ, Buck AA. Comparison of a screening test and a reference test in epidemiologic studies. II. A probabilistic model for the comparison of diagnostic tests. *Am J Epidemiol* 1966;83:593–602.
11. Rogan WJ, Gladen B. Estimating prevalence from the results of a screening test. *Am J Epidemiol* 1978;107:71–6.
12. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Aussagekraft von Surrogatendpunkten in der Onkologie: Rapid report; auftrag A10-05; Version 1.1 [online]. 21.11.2011 [last accessed: 20.11.2019]. (IQWiG-Berichte; Band 80). URL: <https://>



- www.iqwig.de/download/A10-05_Rapid_Report_Version_1-1_Surrogatendpunkte_in_der_Onkologie.pdf.
13. Lachenbruch PA. Sensitivity, specificity, and vaccine efficacy. *Control Clin Trials* 1998;19:569–74.
14. Gyawali B, Hey SP, Kesselheim AS. Evaluating the evidence behind the surrogate measures included in the FDA's table of surrogate endpoints as supporting approval of cancer drugs. *EClinicalMedicine* 2020;21:100332.

