Impact of Case Definitions on Efficacy Estimation in Clinical Trials—A Proof-of-Principle Based on Historical Examples

Andreas Hahn [1] (ID)**, Hagen Frickmann** [1,2,†,*] (ID) **and Andreas E. Zautner** [3,†] (ID)

[1] Institute for Medical Microbiology, Virology and Hygiene, University Medicine Rostock, 18057 Rostock, Germany; hahn.andreas@me.com

[2] Department of Microbiology and Hospital Hygiene, Bundeswehr Hospital Hamburg, 20359 Hamburg, Germany

[3] Institut für Medizinische Mikrobiologie, Universitätsmedizin Göttingen, 37075 Göttingen, Germany; azautne@gwdg.de

* Correspondence: frickmann@bnitm.de; Tel.: +49-40-6947-28743

† These authors contributed equally to this work.

Abstract: Efficacy estimations in clinical trials are based on case definitions. Commonly, they are a more or less complex set of conditions that have to be fulfilled in order to define a clinical case. In the simplest variant, such a case is identical with a single positive diagnostic test result. Frequently, however, case definitions are more complex. Further, their conditions often ignore the inherent logical structure of symptoms and disease: A symptom or a set of symptoms may be necessary but not sufficient for the unambiguous identification of a case. After describing the structure of case definitions and its impact on efficacy estimations, we exemplify this impact using data from two clinical trials dealing with the effectiveness of the vaginal application of tenofovir gel for the prevention of HIV infections and with the therapeutic effects of fecal transplantation on recurrent *Clostridium difficile* infections. We demonstrate that the diagnostic performance of case definitions affects efficacy estimations for interventions in clinical trials. The potential risk of bias and uncertainty is high, irrespective of the complexity of the case definition. Accordingly, case definitions in clinical trials should focus on specificity in order to avoid the risk of bias.

Keywords: efficacy estimation; case definition; specificity; clinical trial

1. Introduction

The endpoints of clinical trials are usually defined by case definitions. Recently, the effect of diagnostic test specificity on the reliability of study outcomes has been demonstrated [1]. In this paper, we show that it is not only test specificity but the combined specificity of both diagnostic tests and clinical case definitions that defines the impact on the study results, making proper handling of both elements, with focus on specificity, advisable prior to conducting a clinical trial.

To demonstrate this hypothesis, we have chosen two well-known historical studies on the specific effects of antimicrobial interventions. The first, the CAPRISA 004 trial, described a double-blind randomized controlled trial on the preventive effects of vaginal tenofovir gel against HIV infection, showing only a moderate effect [2]. With HIV infection as the study endpoint, the case definition is identical with the diagnostic test results, describing the simplest situation. In contrast, the widely referenced study by van Nood and colleagues on the effects of fecal transplantation for the treatment of recurrent *Clostridium difficile* infections [3] was chosen as an example with a slightly more complex and less-specific endpoint. This is because the endpoint, *C. difficile*-associated gastroenteritis, depends

on both the diagnostic detection of the pathogen (i.e., the test component) and the causal attribution of the pathogen to clinical disease (i.e., the case definition component). The latter point is not trivial because, as recently reviewed, determining the etiological relevance of a pathogen in patients with diarrhea may indeed be challenging [4].

Based on these two examples, the effects of the specificity of case definitions are analyzed in a modeling approach.

2. Materials and Methods

If one were to look at the set of all an individual's possible symptoms or attributes, then its power set would be the set of all possible combinations of these symptoms or attributes. A case definition in a very general sense is then any of these combinations:

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

If there is a case definition (e.g., for a disease) and a set of individuals, then 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 Equation (1):

$$Case: \ I \times P(\Omega) \rightarrow \{0, 1\}$$

with

$$Case(i_k, \ S_k) := \begin{cases} 1 \ if \ \exists c \in C : c \subseteq S_k \\ 0 \ else \end{cases} \tag{1}$$

Commonly, the symptoms and attributes of a case definition $C \subseteq P(\Omega)$ are the result of a disease. Then

$$Disease \Rightarrow C \subseteq P(\Omega) \Leftrightarrow Disease \subseteq C \subseteq P(\Omega)$$

holds.

This means that if the symptoms and attributes of a case definition are a result of the disease of interest, then the set of diseased individuals is always a subset of the set of individuals fulfilling the case definition.

Generally, this is not equivalent to the disease itself. In particular,

$$C \subseteq P(\Omega) \Rightarrow Disease \Leftrightarrow C \subseteq P(\Omega) \subseteq Disease$$

does not hold generally. Therefore, a case is not necessarily diseased. Thus, the true disease status is unknown, even in case of a positive case definition.

The true but unknown disease status for every individual $i_k \in I$ is given by Equation (2):

$$Disease \ status: \ I \rightarrow \{0, 1\}$$

with

$$Disease \ status(i_k) := \begin{cases} 1 \ if \ i_k \ diseased \\ 0 \ if \ i_k \ not \ diseased \end{cases} \tag{2}$$

Since the set of individuals fulfilling a case definition of a specific disease and the set of specifically diseased individuals are in general not the same, a case definition can be understood as a diagnostic test for its disease.

Then, the sensitivity of a case definition is given by Equation (3):

$$Se : I \rightarrow [0, 1]$$

with

$$\frac{\left|\{i_k : Disease\ status\ (i_k) = 1\} \cap \{i_k :\ Case\ (i_k) = 1\}\right|}{\left|\{i_k : Disease\ status\ (i_k) = 1\}\right|}$$ (3)

In the same way, the specificity of a case definition is given by Equation (4):

$$Sp : I \rightarrow [0, 1]$$

with

$$\frac{\left|\{i_k : Disease\ status\ (i_k) = 0\} \cap \{i_k :\ Case\ (i_k) = 0\}\right|}{\left|\{i_k : Disease\ status\ (i_k) = 0\}\right|}$$ (4)

Obviously, the definitions above lead to the known definitions for sensitivity and specificity:

$$Se\ (Case\ definition) = P(Case = 1 | Disease\ status = 1)$$

and

$$Sp\ (Case\ definition) = P(Case = 0 | Disease\ status = 0)$$

A case definition C is perfectly sensitive if the set of diseased individuals is a subset of all cases:

$$\{i_k : Disease\ status\ (i_k) = 1\} \subseteq \{i_k : Case(i_k) = 1\}$$

This is the same as:

$$Disease \Rightarrow C \subseteq P(\Omega)$$

A case definition C is completely specific if the set of non-diseased individuals is a subset of all non-cases:

$$\{i_k : Disease\ status\ (i_k) = 0\} \subseteq \{i_k : Case(i_k) = 0\}$$

If a case definition is completely specific, then we have:

$$C \subseteq P(\Omega) \Rightarrow Disease$$

Only if a case definition is completely sensitive as well as completely specific is it equivalent to the disease:

$$C \subseteq P(\Omega) \Leftrightarrow Disease$$

This means that the set of diseased individuals is identical with the set of individuals fulfilling the case definition.

Since every set of symptoms or attributes should be verified by a diagnostic test system that has its own sensitivity and specificity as a combination of all individual sensitivities and specificities of every component of the diagnostic test system, our initial definition of a case has to be modified as follows:

Let I be the set of all individuals. $S_k \in P(\Omega)$ denotes the individual set of symptoms and attributes for individual $i_k \in I$. Let the diagnostic test system be a map from the power set of symptoms and attributes to the tuple (0,1), where 0 indicates a negative and 1 indicates a positive diagnostic test result:

$$Diagnostic\ test\ system :\ P(\Omega) \rightarrow \{0, 1\}$$

A case is defined by Equation (5):

$$Case :\ I \times P(\Omega) \rightarrow \{0, 1\}$$

with

$$Case(i_k,\ S_k) := \begin{cases} 1\ if\ \exists c \in C : c\ \subseteq\ S_k\ with\ Diagnostic\ test\ system(c) = 1 \\ 0\ else \end{cases}$$ (5)

This means that a case is an individual for whom there exists a set of symptoms or attributes of the case definition for whom the diagnostic test system conducted will be positive.

The impacts of diagnostic sensitivity and specificity on prevalence or intervention efficacy estimations were described by Gart and Buck [5,6] and Rogan and Gladen [7] as well as Lachenbruch [8], Gart [9], and Neyman [10]. They proposed sensitivity- and specificity-adjusted point estimators for prevalence or incidence and intervention efficacy. While the authors above focused specifically on diagnostic tests, we now point out that even a case definition itself has its own sensitivity and specificity regarding the disease for which the case definition has been defined. Consequently, the overall sensitivity and specificity of a case definition is always a combination of the sensitivity and specificity of the test system testing for the attributes and symptoms of the case definition, as well as the sensitivity and specificity of the case definition itself.

If overall sensitivity and specificity are known, then it is possible to apply the sensitivity- and specificity-adjusted point estimators of the authors quoted above.

Given a case definition C of a disease with sensitivity Se and specificity Sp, then the sensitivity- and specificity-adjusted prevalence estimator $Prev$ is given by Equation (6):

$$Prev = \frac{\frac{\sum_{k=1}^{n} Case(i_k, S_k)}{n} - 1 + Sp}{Se + Sp - 1} \tag{6}$$

Based on the estimator above in (6), a sensitivity- and specificity-adjusted prevention efficacy PE estimator is given by Equation (7):

$$PE = 1 - \frac{\frac{\sum_{k=1}^{n} Case(i_k, S_k)}{n} - 1 + Sp}{\frac{\sum_{l=1}^{n} Case(i_l, S_l)}{n} - 1 + Sp} \tag{7}$$

The estimator above in (7) holds for the specificity of a case definition that is constant over the study arms. Obviously, specificity of the case definition is fundamental for the bias in an unadjusted estimation that uses the raw rate of cases in both arms for efficacy estimation. This explains why case definitions in clinical trials may differ from case definitions in the clinical context. Clinical trials are designed to estimate unbiased effects in a population, while the focus of a clinical case definition is the individual patient.

Case definitions should be designed to be specific, considering their focus in clinical trials. Furthermore, variation of specificity over the study arms should be avoided. Accordingly, an open-label design in accordance with investigator-assessed endpoints or patient-reported outcomes should be avoided.

For cases where the sensitivity or specificity of a case definition vary over study arms, an adjusted estimator is given by Equation (8):

$$\frac{\left(\frac{\sum_{k=1}^{n} Case(i_{1k}, S_{1k})}{n} - 1 + Sp_1\right)(Se_2 + Sp_2 - 1)}{\left(\frac{\sum_{l=1}^{n} Case(i_{2l}, S_{2l})}{n} - 1 + Sp\right)(Se_1 + Sp_1 - 1)} \tag{8}$$

The 0.95 confidence intervals are given by Equation (9) as follows:

$$0.95\ CI = 1 - e^{Ln(1-IE) \pm 1.96\left(\frac{1}{prev_1^2} Var(Prev_1) + \frac{1}{prev_2^2} Var(Prev_2)\right)}$$

with

$$Var(Prev) = \frac{\frac{\sum_{k=1}^{n} Case(i_k, S_k)}{n}\left(1 - \frac{\sum_{k=1}^{n} Case(i_k, S_k)}{n}\right)}{n\ (Se + Sp - 1)^2} \tag{9}$$

One of the main obstacles for applying the sensitivity- and specificity-adjusted estimators above is that the diagnostic sensitivity and specificity have to be known. Since the true disease status is likely unknown, indirect methods are often used for the evaluation of diagnostic sensitivity and specificity. Methods and problems regarding the assumptions of such approaches have been widely discussed [10–20].

The validity of indirect methods for the estimation of the diagnostic sensitivity and specificity thus depends on the fulfillment of the underlying model assumptions. Ideally, the case definition should be defined with a specificity virtually equal to 1, and it should not vary over treatment arms, leading to a minimum of bias as a consequence of Equation (7).

2.1. Summaries of Historical Studies Used to Exemplify the Approach

2.1.1. The CAPRISA 004 Trial

The case definitions of the CAPRISA 004 trial on the preventive effects of vaginal tenofovir gel against HIV infection [2] and, in comparison, the study by van Nood and colleagues on the effects of fecal transplantation for the treatment of recurrent *Clostridium difficile* infections [3] are analyzed and discussed.

The double-blind, randomized, controlled CAPRISA 004 trial assessed effectiveness and safety of a 1% vaginal gel formulation of tenofovir for preventing the acquisition of HIV. It was conducted to compare tenofovir gel ($n = 445$) with placebo gel ($n = 444$) in sexually active, non-HIV-infected 18–40-year-old women in urban and rural KwaZulu-Natal, South Africa. At monthly follow-up visits for 30 months, the parameters HIV serostatus, safety, sexual behavior, and gel and condom use were assessed. In the tenofovir gel arm, reported HIV incidence was 5.6 per 100 woman-years (wy, 38/680.6 wy), compared to 9.1 per 100 wy (60/660.7 wy) in the placebo arm. The overall protective efficacy against HIV infection was estimated at 39%. Two HIV rapid tests, Determine® HIV-1/2 (Abbott Laboratories, IL, USA) and Uni-Gold Recombigen® HIV test (Trinity Biotech, Wicklow, Ireland), were applied during each study visit. By protocol, only HIV infections during study follow-up in eligibly enrolled women, as confirmed by two independent RNA PCR results, were defined as study endpoints. Participants in the HIV window period at the end of the study were included as HIV-related endpoints if seropositivity was confirmed after the study.

2.1.2. The Study by van Nood and Colleagues

The study by van Nood and colleagues investigated the effect of duodenal infusion of donor feces in patients with recurrent *C. difficile* infection. The study patients were randomly assigned to receive one of three therapeutic approaches:

- An initial vancomycin regimen of 500 mg orally four times per day for 4 days, followed by bowel lavage and subsequent infusion of a solution of donor feces through a nasoduodenal tube;
- A standard vancomycin regimen of 500 mg orally four times per day for 14 days; or
- A standard vancomycin regimen with bowel lavage.

The primary endpoint was the resolution of diarrhea associated with *C. difficile* infection without relapse after 10 weeks. For this purpose, "resolution of diarrhea associated with *C. difficile* infection" was defined as the absence of diarrhea or persistent diarrhea that could be explained by other causes with three consecutive negative stool tests for *C. difficile* toxin.

3. Results and Discussion

In the following we demonstrate the effect of case definition for the two historical clinical trials described above.

3.1. Effectiveness and Safety of Tenofovir Gel, an Antiretroviral Microbicide, for the Prevention of HIV Infection in Women

The structure of the case definition in this study is simple since there is only one attribute to be confirmed—An HIV infection after inclusion into the study. Thus, a case according to protocol (see also Equation (10)) was defined as a woman who was HIV-negative at the time of inclusion in the study with four positive test results indicating HIV infection (two positive HIV rapid tests and two positive RNA PCR results) in the course of the study:

$$Case: \ I \times P(\Omega) \rightarrow \{0, 1\}$$

with

$$Case(i_k, \ S_k) := \begin{cases} 1 \ if \ all \ 4 \ conducted \ HIV \ tests \ of \ the \ individuum \ i_k \ are \ positive \\ 0 \ else \end{cases} \tag{10}$$

As such, the overall sensitivity and specificity are given by the diagnostic test sensitivity and specificity of the four diagnostic tests according to Equations (11) and (12):

$$Sensitivity_{overall} = \prod_{i=1}^{4} Sensitivity_i \tag{11}$$

$$Specificity_{overall} = 1 - \prod_{i=1}^{4} (1 - Specificity_i) \tag{12}$$

For the Determine® HIV-1/2 rapid test (Abbott Laboratories), sensitivity of 0.998 and specificity of 0.994, and for the Uni-Gold Recombigen® HIV test (Trinity Biotech), sensitivity of 0.985 and specificity of 0.995 were reported by Piwowar-Manning and colleagues [21]. The diagnostic performance of the PCRs conducted is not reported. Nevertheless, it can be assumed that the sensitivity and specificity of these PCRs are very high. Assuming that the PCR specificity is 0.99 or higher, the overall specificity of the case definition is practically 1. Using Equation (7) from the Materials and Methods section, it follows that the case definition leads to a practically unbiased efficacy estimation even without any adjustment (see also Equation (13)):

$$PE = 1 - \frac{\frac{\sum_{k=1}^{n} Case(i_k, S_k)}{n} - 1 + Sp}{\frac{\sum_{l=1}^{n} Case(i_l, S_l)}{n} - 1 + Sp} = 1 - \frac{\frac{\sum_{k=1}^{n} Case(i_k, S_k)}{n} - 1 + 1}{\frac{\sum_{l=1}^{n} Case(i_l, S_l)}{n} - 1 + 1} = 1 - \frac{\frac{\sum_{k=1}^{n} Case(i_k, S_k)}{n}}{\frac{\sum_{l=1}^{n} Case(i_l, S_l)}{n}} \tag{13}$$

Nevertheless, in Table 1 and Figure 1, we illustrate the variation of adjusted efficacy estimates with 0.95 confidence intervals calculated by Equation (9) for varying specificity assumptions. The unadjusted estimation is identical with the case of ideal diagnostic specificity that can be assumed for this trial. Therefore, unadjusted and adjusted estimations are the same if a diagnostic specificity of 1 can be assumed.

Table 1. HIV prevention efficacy estimates depending on specificity assumptions using Equations (7) and (9).

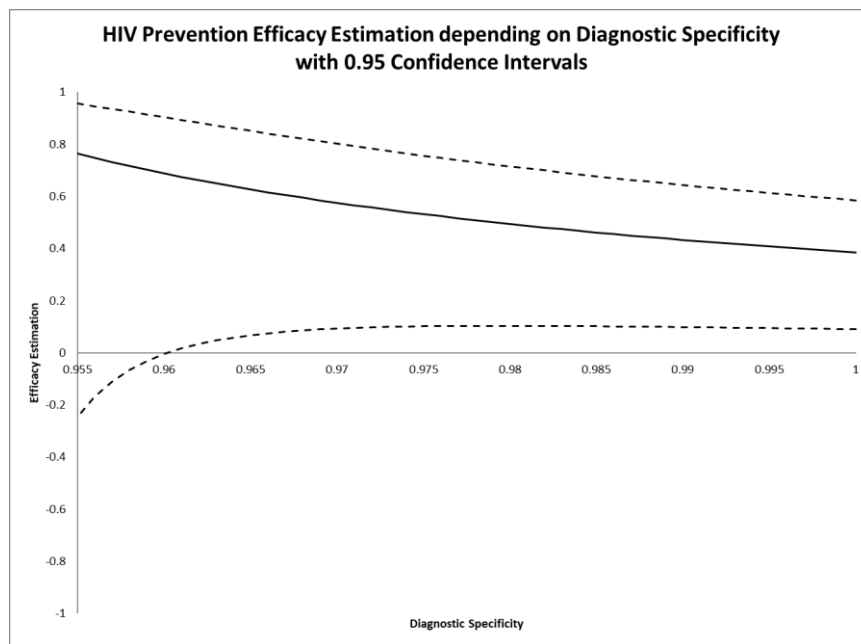| Specificity Assumption	Efficacy Estimation	0.95 Confidence Interval
1	0.385	0.090, 0.584
0.995	0.408	0.094, 0.613
0.985	0.461	0.101, 0.677
0.975	0.532	0.102, 0.756
0.965	0.627	0.067, 0.851
0.955	0.764	−0.246, 0.955

Figure 1. Adjusted efficacy estimates (*y*-axis) depending on endpoint specificity (*x*-axis) applying Equation (7) with 0.95 confidence intervals and using Equation (9) for the primary endpoint of the CAPRISA 004 trial.

Although CAPRISA 004 was among the least-efficient HIV pre-exposure prophylaxis (PrEP) trials [22,23] for various reasons, with a major limitation in the field of adherence [24,25], the case definition was very well chosen, and later trials confirmed the preventive usefulness of the HIV PrEP approach with similar definitions of the clinical endpoint [26,27].

3.2. Duodenal Infusion of Donor Feces for Recurrent Clostridium difficile

The case definition based on the endpoint "resolution of diarrhea associated with *C. difficile* infection," meaning an individual who was initially cured had a relapse within 10 weeks after initiation of therapy, can be reduced to "diarrhea" and "at least one positive test out of three tests for *C. difficile* toxin." (see also Equation (14))

$$Case: \ I \times P(\Omega) \to \{0, 1\}$$

with

$$Case(i_k, S_k) := \begin{cases} 1 \ if \ i_k \ has \ diarrhea \ and \ at \ least \ 1 \ of \ 3 \ conducted \ tests \\ \qquad\qquad for \ C.difficile \ is \ positive \\ \qquad\qquad\qquad 0 \ else \end{cases} \tag{14}$$

Assuming that the diagnosis of diarrhea as a clinical symptom is always true, the overall sensitivity and specificity of the case definition depend on the test system for *C. difficile* toxin. Its sensitivity and specificity (see also Equations (15) and (16)) are given by:

$$Sensitivity_{overall} = 1 - \prod_{i=1}^{3} (1 - Sensitivity_i) \tag{15}$$

$$Specificity_{overall} = \prod_{i=1}^{3} Specificity_i \tag{16}$$

The reported sensitivity and specificity for the test conducted (Premier Toxin A and B Assay) were 0.9744 and 0.9752 as reported by Novak-Weekley and colleagues [28]. This leads to an overall endpoint specificity of 0.927 and an overall endpoint sensitivity of 0.99998. Compared with the

CAPRISA 004 trial, the case definition of this trial leads to a maximum endpoint sensitivity, while the endpoint specificity is reduced. As follows from Equation (7) from the Materials and Methods section, the reduced specificity leads to a bias in efficacy estimation. As presented in Table 2 and Figure 2, where adjusted efficacy estimations with 0.95 confidence intervals are given for varying endpoint specificity, the overall specificity of 0.927 leads to a loss of significance for the adjusted efficacy estimation. Contrary to the HIV trial modeled above, the case definition in this trial does not result in a maximum specificity that would lead to an unbiased efficacy estimation even without adjustment. A review of controversies associated with therapeutic fecal transplantation was published by van Nood and colleagues one year after the publication of the study addressed in this modeling [29]. The allocation of etiological relevance to pathogens in human samples remains an issue of ongoing debate [30–32].

Table 2. *Clostridium difficile* relapse prevention efficacy estimates depending on specificity assumptions using Equations (7) and (9).

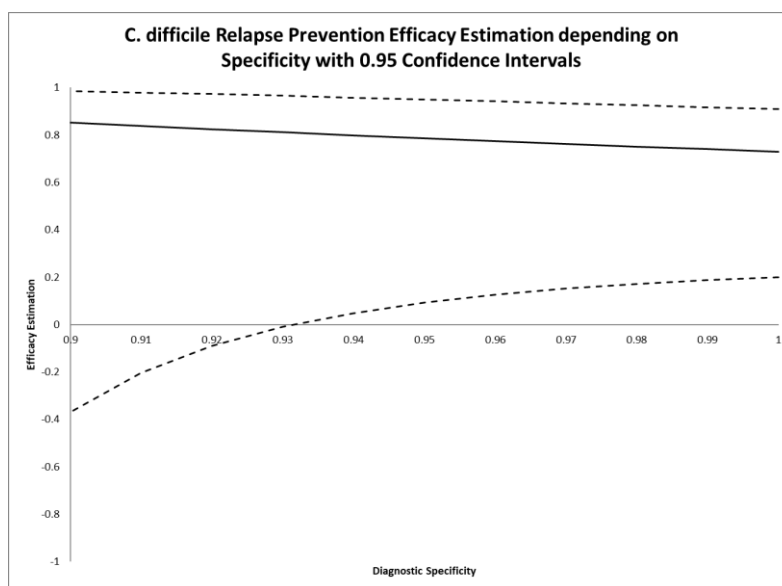| Specificity Assumption	Efficacy Estimation	0.95 Confidence Interval
1	0.729	0.201, 0.908
0.98	0.751	0.172, 0.925
0.96	0.774	0.126, 0.942
0.94	0.798	0.048, 0.957
0.927	0.815	−0.031, 0.967
0.92	0.824	−0.090, 0.972
0.90	0.852	−0.370, 0.984

Figure 2. Adjusted efficacy estimates (*y*-axis) depending on endpoint specificity (*x*-axis) applying Equation (7) with 0.95 confidence intervals and using Equation (9) for the primary endpoint of the study by van Nood and colleagues on the effects of fecal transplantation for the treatment of recurrent *Clostridium difficile* infections.

As mentioned above, case definitions can be described as subsets of the power set of the set of all possible human attributes including clinical symptoms and laboratory parameters. Since clinical symptoms especially are a result of a disease and not the disease itself, the inference of a disease from a symptom is logically problematic. It follows that case definitions have their own sensitivity and specificity, just as classical diagnostic methods do [1]. From Lachenbruch's sensitivity- and specificity-adjusted efficacy estimator [8], it also follows that the diagnostic specificity in particular is critical for the bias of an unadjusted estimation.

Since varying sensitivity and specificity over study arms makes the situation much more complex, the study design is also very critical. For example, open-label studies with subjective attributes such as patient-reported outcomes as part of the case definition could result in different diagnostic performance over the study arms.

Further, and as demonstrated by the examples, it should be noted that diagnostic performance of study endpoints is always a combination of case definition as a set of attributes and diagnostic tests verifying these attributes. The structures of case definitions in the trials presented were fundamentally different: The case definition of the first trial led to a very specific endpoint definition resulting in an unbiased efficacy estimation. The case definition of the second trial maximized the endpoint sensitivity and clearly reduced the endpoint specificity, necessarily resulting in biased efficacy estimation with a possible loss of statistical significance. Therefore, case definitions in clinical trials should focus on endpoint specificity in order to avoid bias or the possible necessity of adjusting for endpoint sensitivity and specificity requiring the evaluation of diagnostic sensitivity and specificity, which is a complex procedure by itself. To accomplish this, the diagnostic systems applied have to be taken into account in order to not lose overall specificity as a result of applying an unspecific diagnostic approach. The consequence could be differing case definitions depending on their use either in clinical practice or in clinical trials.

References

1. Hahn, A.; Meyer, C.G.; Frickmann, H. Impact of diagnostic methods on efficacy estimation—A proof-of-principle based on historical examples. *Trop. Med. Int. Health* **2020**, *25*, 357–363. [CrossRef] [PubMed]
2. Abdool Karim, Q.; Abdool Karim, S.S.; Frohlich, J.A.; Grobler, A.C.; Baxter, C.; Mansoor, L.E.; Kharsany, A.B.; Sibeko, S.; Mlisana, K.P.; Omar, Z.; et al. Effectiveness and safety of tenofovir gel, an antiretroviral microbicide, for the prevention of HIV infection in women. *Science* **2010**, *329*, 1168–1174. [CrossRef]
3. van Nood, E.; Vrieze, A.; Nieuwdorp, M.; Fuentes, S.; Zoetendal, E.G.; de Vos, W.M.; Visser, C.E.; Kuijper, E.J.; Bartelsman, J.F.; Tijssen, J.G.; et al. Duodenal infusion of donor feces for recurrent *Clostridium difficile*. *N. Engl. J. Med.* **2013**, *368*, 407–415. [CrossRef] [PubMed]
4. Zautner, A.E.; Groß, U.; Emele, M.F.; Hagen, R.M.; Frickmann, H. More pathogenicity or just more pathogens?—On the interpretation problem of multiple pathogen detections with diagnostic multiplex assays. *Front. Microbiol.* **2017**, *8*, 1210. [CrossRef] [PubMed]
5. Gart, J.J.; Buck, A.A. 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. [CrossRef]
6. Buck, A.A.; Gart, J.J. Comparison of a screening test and a reference test in epidemiologic studies. I. Indices of agreement and their relation to prevalence. *Am. J. Epidemiol.* **1966**, *83*, 586–592. [CrossRef]
7. Rogan, W.J.; Gladen, B. Estimating prevalence from the results of a screening test. *Am. J. Epidemiol.* **1978**, *107*, 71–76. [CrossRef]
8. Lachenbruch, P.A. Sensitivity, specificity, and vaccine efficacy. *Control Clin. Trials.* **1998**, *19*, 569–574. [CrossRef]
9. Gart, J.J. The estimation problem in a probabilistic model for the comparison of diagnostic tests (Abstract). *Biometrics* **1965**, *SJ*, 764.

10. Neyman, J. Outline of statistical treatment of the problem of diagnosis. *Pub. Health Rep.* **1947**, *63*, 1449–1456. [CrossRef]

11. Hui, S.L.; Walter, S.D. Estimating the error rates of diagnostic tests. *Biometrics* **1980**, *36*, 167–171. [CrossRef]

12. Hui, S.L.; Zhou, X.H. Evaluation of diagnostic tests without gold standards. *Statist. Meth. Med. Res.* **1998**, *7*, 354–370. [CrossRef] [PubMed]

13. Hadgu, A.; Qu, Y. A biomedical application of latent class models with random effects. *Appl. Statist.* **1998**, *47*, 603–616. [CrossRef]

14. Lu, D.; Zhou, C.; Tang, L.; Tan, M.; Yuan, A.; Chan, L. Evaluating accuracy of diagnostic tests without conditional independence assumption. *Stat. Med.* **2018**, *37*, 2809–2821. [CrossRef] [PubMed]

15. Torrance-Rynard, V.L.; Walter, S.D. Effects of dependent errors in the assessment of diagnostic test performance. *Statist. Med.* **1997**, *16*, 2157–2175. [CrossRef]

16. Weng, T.S. Evaluation of a new diagnostic test against a reference test less than perfect in accuracy. *Commun. Statist. Simul. Comput.* **1996**, *35*, 533–555. [CrossRef]

17. Vacek, P.M. The effect of conditional dependence on the evaluation of diagnostic tests. *Biometrics* **1985**, *41*, 959–968. [CrossRef]

18. Valenstein, P.N. Evaluating diagnostic tests with imperfect standards. *Am. J. Clin. Pathol.* **1990**, *93*, 252–258. [CrossRef]

19. Enøe, C.; Andersen, S.; Thomsen, L.K.; Mousing, J.; Leontides, L.; Sùrensen, V.; Willeberg, P. Estimation of the sensitivity and the specificity of two diagnostic tests for the detection of antibodies against *Actinobacillus pleuropneumoniae* serotype 2 in pigs by maximum-likelihood-estimation and Gibbs sampling. *Epidémiologie St. Anim.* **1997**, *31/32*, 12.

20. Enøe, C.; Georgiadis, M.P.; Johnson, W.O. Estimation of sensitivity and specificity of diagnostic tests and disease prevalence when the true disease state is unknown. *Prev. Vet. Med.* **2000**, *45*, 61–81. [CrossRef]

21. Piwowar-Manning, E.M.; Tustin, N.B.; Sikateyo, P.; Kamwendo, D.; Chipungu, C.; Maharaj, R.; Mushanyu, J.; Richardson, B.A.; Hillier, S.; Brooks Jackson, J. Validation of rapid HIV antibody tests in 5 African countries. *J. Int. Assoc. Physicians AIDS Care* **2010**, *9*, 170–172. [CrossRef] [PubMed]

22. Karim, S.S.; Karim, Q.A. Antiretroviral prophylaxis: A defining moment in HIV control. *Lancet* **2011**, *378*, e23–e25. [CrossRef]

23. Valley-Omar, Z.; Sibeko, S.; Anderson, J.; Goodier, S.; Werner, L.; Arney, L.; Naranbhai, V.; Treurnicht, F.; Abrahams, M.R.; Bandawe, G.; et al. CAPRISA 004 tenofovir microbicide trial: No impact of tenofovir gel on the HIV transmission bottleneck. *J. Infect. Dis.* **2012**, *206*, 35–40. [CrossRef]

24. MacQueen, K.M.; Weaver, M.A.; van Loggerenberg, F.; Succop, S.; Majola, N.; Taylor, D.; Karim, Q.A.; Karim, S.A. Assessing adherence in the CAPRISA 004 tenofovir gel HIV prevention trial: Results of a nested case-control study. *AIDS Behav.* **2014**, *18*, 826–832. [CrossRef] [PubMed]

25. Mansoor, L.E.; Abdool Karim, Q.; Yende-Zuma, N.; MacQueen, K.M.; Baxter, C.; Madlala, B.T.; Grobler, A.; Abdool Karim, S.S. Adherence in the CAPRISA 004 tenofovir gel microbicide trial. *AIDS Behav.* **2014**, *18*, 811–819. [CrossRef] [PubMed]

26. McCormack, S.; Dunn, D.T.; Desai, M.; Dolling, D.I.; Gafos, M.; Gilson, R.; Sullivan, A.K.; Clarke, A.; Reeves, I.; Schembri, G.; et al. Pre-exposure prophylaxis to prevent the acquisition of HIV-1 infection (PROUD): Effectiveness results from the pilot phase of a pragmatic open-label randomised trial. *Lancet* **2016**, *387*, 53–60. [CrossRef]

27. Molina, J.M.; Capitant, C.; Spire, B.; Pialoux, G.; Cotte, L.; Charreau, I.; Tremblay, C.; Le Gall, J.M.; Cua, E.; Pasquet, A.; et al. On-demand preexposure prophylaxis in men at high risk for HIV-1 infection. *N. Engl. J. Med.* **2015**, *373*, 2237–2246. [CrossRef]

28. Novak-Weekley, S.M.; Hollingsworth, M.H. Comparison of the premier toxin A and B assay and the TOX A/B II assay for diagnosis of *Clostridium difficile* infection. *Clin. Vaccine Immunol.* **2008**, *15*, 575–578. [CrossRef]

29. van Nood, E.; Speelman, P.; Nieuwdorp, M.; Keller, J. Fecal microbiota transplantation: Facts and controversies. *Curr. Opin. Gastroenterol.* **2014**, *30*, 34–39. [CrossRef]

30. Lindsay, B.; Ochieng, J.B.; Ikumapayi, U.N.; Toure, A.; Ahmed, D.; Li, S.; Panchalingam, S.; Levine, M.M.; Kotloff, K.; Rasko, D.A.; et al. Quantitative PCR for detection of *Shigella* improves ascertainment of *Shigella* burden in children with moderate-to-severe diarrhea in low-income countries. *J. Clin. Microbiol.* **2013**, *51*, 1740–1746. [CrossRef]

31.   Liu, J.; Kabir, F.; Manneh, J.; Lertsethtakarn, P.; Begum, S.; Gratz, J.; Becker, S.M.; Operario, D.J.; Taniuchi, M.; Janaki, L.; et al. Development and assessment of molecular diagnostic tests for 15 enteropathogens causing childhood diarrhoea: A multicentre study. *Lancet Infect. Dis.* **2014**, *14*, 716–724. [CrossRef]

32.   Platts-Mills, J.A.; Gratz, J.; Mduma, E.; Svensen, E.; Amour, C.; Liu, J.; Maro, A.; Saidi, Q.; Swai, N.; Kumburu, H.; et al. Association between stool enteropathogen quantity and disease in Tanzanian children using TaqMan array cards: A nested case—control study. *Am. J. Trop. Med. Hyg.* **2014**, *90*, 133–138. [CrossRef] [PubMed]