Published in final edited form as:

Curr Treat Options Infect Dis. 2014 July 29; 6(3): 326–336. doi:10.1007/s40506-014-0026-0.

Outcome Assessments in Clinical Trials of Cryptococcal Meningitis: Considerations on Use of Early Fungicidal Activity as a Potential Surrogate Endpoint for All-Cause Mortality

Jairo Mauricio Montezuma-Rusca, MD and John H. Powers, MD*

Opinion statement

Cryptococcal meningitis (CM) is a common disease in resource-challenged settings, with a high mortality within weeks of disease onset. Mortality remains high with current treatments, so more effective interventions are needed to decrease mortality. There has been interest in using the outcome assessment of quantification of fungus from cerebrospinal fluid as a replacement (surrogate) endpoint for all-cause mortality (ACM) as a means of decreasing sample size in randomized clinical trials in CM. To evaluate a biomarker as a potential surrogate endpoint to replace ACM requires several steps. This paper discusses the issues of determining whether the context of a disease is one where a potential surrogate endpoint is rational, the types of outcome assessments that might qualify as potential surrogates, and the process for evaluation of the evidence that a chosen biomarker is a valid replacement for ACM in the given context of use. We then apply those principles to the context of randomized clinical trials of CM.

Keywords

Surrogate endpoints; Biomarkers; Cryptococcus; Cryptococcal meningitis; Clinical trials; Meningitis; Mortality; All-cause mortality; Early fungicidal activity; Fungal burden

Introduction

Cryptococcal meningitis (CM) is a major cause of mortality in patients with HIV/AIDS, particularly in populations where combined antiretroviral therapy is not readily available. There are approximately 1 million cases per year and 625,000 deaths. In sub-Saharan Africa, it is the leading cause of death among people living with HIV/AIDS, with case fatality rates as high as 70 % in 3 months [1–3]. Although antifungal treatments for CM, including triazole and amphotericin-based therapy, have been studied, high mortality rates are still reported in treated patients in recent trials [4]. Therefore, more effective therapies are needed. Adequately designed clinical trials are necessary to evaluate the benefits and harms of interventions in the treatment of CM.

*Address: 6700B Rockledge Drive Room 1123, Bethesda, MD 20817, USA, John.powers@nih.gov.

Conflict of Interest

John H. Powers and Jairo Mauricio Montezuma-Rusca declare no conflicts of interest.

Compliance with Ethics Guidelines

Human and Animal Rights and Informed Consent

This article does not contain any studies with human or animal subjects performed by any of the authors.

In designing trials to evaluate interventions in the treatment of CM, interest has increased regarding the selection of outcome assessments as endpoints, including biomarkers as potential surrogate endpoints for mortality. The purpose of these biomarkers would be to use a smaller sample size in randomized trials to determine which treatment options for CM might lower mortality compared with other treatment regimens. We describe considerations in the selection of outcome assessments in clinical trials in general, and then apply those principles to the setting of clinical trials in CM. We also discuss the contexts in which selection of biomarkers as surrogate endpoints is scientifically justifiable and the evaluation of the validity of biomarkers as surrogate endpoints.

Outcome assessments in clinical trials

The primary goal of administering any intervention is to have beneficial net effects (balanced against harms) on meaningful health aspects of patients' lives, defined as patient symptoms, daily function, and/or survival. The selection of appropriate outcome assessments for use as endpoints in clinical trials depends on the context of use of the trial. Context of use includes the natural history of the disease, the patient population under study, the severity of illness at baseline in that population, and the study design, among other considerations.

Outcome assessments in serious and life-threatening diseases

For serious and life-threatening diseases, the most important meaningful health aspect for which treatment should have an effect is in decreasing all-cause mortality (ACM) and irreversible morbidity for patients. In acute diseases, ACM can be measured in a short period of time after randomization. ACM is of primary importance, but ACM also obviates the measurement of other outcomes within the same time period; that is, symptoms such as neurological abnormalities cannot be measured in patients who have died during the course of a study. ACM is more clinically relevant than cause-specific mortality. In a population of patients who die of a disease in the given time period, treatments for a different disease cannot be helpful to those patients in the same time period. For instance, if patients with CM die of tuberculosis or HIV during the time of treatment for CM, those other causes are not irrelevant; rather, those patients do not benefit from treatment for CM. Cause-specific mortality attempts to answer a less clinically relevant question of whether patients would live if they did not die of something else first. ACM as an outcome assessment is objective in that it does not require interpretation from others, as opposed to cause-specific mortality, which often requires judgment in ascertaining a cause of death. Studies show that clinicians often are inaccurate in their assessments of cause of death compared with autopsy data [5]. The process of randomization in clinical trials gives an equal probability of death from causes other than the interventions tested. Randomization allows for conclusions that differences in outcomes – either for benefit or for harm – are causally related to the interventions tested [6]. ACM does not require any special testing to measure, and data on ACM usually can be obtained even in resource-challenged settings. Finally, ACM encompasses both benefit and harm into a single measure. Interventions that have beneficial effects on one aspect of disease may be counterbalanced if the intervention increases

mortality through some other unintended off-target mechanism of disease or through adverse effects of the intervention [7, 8•].

Outcome assessments that measure physiological or pharmacological processes measure biological activity but do not necessarily reflect benefits on meaningful health aspects of patients' lives, such as symptoms, function, or survival [9]. While biological activity may confirm a mechanism of action or help to select candidate interventions for further study in confirmatory trials, biological activity is insufficient to evaluate whether patient benefit exceeds harm on patient-centered outcomes. In all clinical settings, and especially in resource-challenged settings, demonstration of patient benefit that exceeds harm justifies paying for and implementing interventions in clinical practice.

Direct and indirect outcome assessments

Outcome assessments that directly measure patients' symptoms, function, and survival have been termed 'clinical' outcomes [9]. This term does cause confusion since the word 'clinical' is often used in other contexts to refer to any measurement made at the patient's bedside [8•]. However, many indirect measures of benefit are measured at the bedside in the clinical setting. Indirect measures of patient benefit include clinician-reported outcomes of signs of disease that require some judgment of trained medical professionals in obtaining or interpreting the measurement, and laboratory or radiological measures that do not require judgment in their measurement or interpretation. The latter are defined as biomarkers [9]. Biomarkers can serve a number of purposes, including diagnosis of disease, staging of disease, as risk factors for acquisition of disease, and as prognostic measures that indicate which patients are more likely to have worse outcomes independent of treatment [9]. In some settings, biomarkers have been proposed as outcome assessments to define endpoints to measure treatment effects of interventions in clinical trials. In this context, they function as replacements or 'surrogates' for the actual direct outcome of interest on patient symptoms, function, or survival.

Outcome assessments are the tools/measures used to evaluate health status of patients, while an endpoint is how the outcome assessment is used in the setting of clinical trials to measure treatment effects [10]. For example, microbiological biomarkers based on cultures of organisms are outcome assessments. However, this biomarker outcome assessment of cultures can be evaluated in various ways to form endpoints in trials. For instance, cultures can be evaluated as a dichotomous value (culture positive/culture negative) at a fixed time-point, as time to culture conversion, or as the slope of decrease in culture positivity. Results based on these endpoints can be evaluated using mean differences between groups or developing a 'responder' criteria based on proportions of patients in the study group who meet some defined 'successful' outcome based on cultures. All use the same outcome assessment (microbiological culture), but the endpoint is constructed in different ways. Endpoints can be composed of more than one outcome assessment. For instance, when biomarkers are used as surrogate endpoints in trials, any events for which the biomarker is a replacement (i.e. ACM) that occur during the course of the study would still be considered 'failure' of treatment. For example, in trials of antiretroviral agents for HIV/AIDS, any

deaths that occur during a study would count as ‘failure’ of treatment, regardless of viral load measurements.

Choosing biomarkers as indirect outcome assessments

There are several considerations regarding the choice of biomarkers as indirect measures of patient benefit in clinical trials. The first is to decide on the meaningful health aspects of patients’ lives based on survival or patient function/symptoms for which the biomarker may serve as a replacement. Lack of clarity on this basic point can lead to challenges in both evaluating the biomarker and in interpreting trial results. In most situations, the not insubstantial risks of erroneous results with using a biomarker as a surrogate endpoint are justifiable only in the setting of serious and life-threatening diseases, meaning the biomarker serves as substitute for ACM or some direct measure of irreversible morbidity. US regulations point out that surrogate endpoints as a basis for accelerated approval of drugs and biologics are only used in the setting of serious and life-threatening diseases [11]. Part of the reason for this is that when biomarkers serve as substitutes for irreversible morbidity and mortality, there is justification for a benefit that outweighs any non-mortality harms. This is because when biomarkers are used to measure benefit, harms and benefit may be judged on different scales, since harms are usually evaluated through direct outcome assessments. Therefore, the comparison of benefit and harms is challenging when benefit is measured based on a biomarker, but harm is measured based on direct assessments, for example comparing a serious adverse event like anaphylaxis to a laboratory-based indirect assessment of benefit like microbiological cultures. This occurred in a recent example with the drug bedaquiline for multi-drug-resistant tuberculosis, where the drug decreased sputum culture positivity when added to older TB drugs compared with those same older TB drugs as the control, but increased mortality fivefold [12].

A second consideration is to evaluate whether a biomarker as a potential surrogate endpoint is needed in the specific context of use of a clinical trial due to challenges in measuring a direct outcome. Biomarkers most often are proposed as outcomes in chronic diseases where the direct outcome of interest like ACM may happen years in the future. Biomarkers are less often needed in acute diseases where direct outcomes occur in weeks to months. In the setting of acute diseases, biomarkers may be used if the direct outcomes of interest are uncommon. Since biomarkers may be more common and treatment effects may be larger, they may allow for smaller sample sizes for trials [8•].

If the context of use justifies use of a biomarker as a substitute for a clearly defined direct outcome, the next step is to choose a candidate biomarker. Candidate biomarkers should be easily measured in a reproducible manner. Biomarkers should be on the causal pathway of the disease. Biomarkers that are not on the causal pathway of disease may be affected independently of the effect on the disease. For instance, increased body temperature (fever) is not on the causal pathway of infection, rather infection causes fever. Therefore, body temperature may be lowered by interventions (e.g. antipyretics) that do not affect the course of infection.

Evaluating biomarkers: relationship to all-cause mortality

The treatment effect of an intervention on a biomarker should reflect the net treatment effect of the intervention on benefits and harms on direct measures of benefit such as ACM.

Treatment effect is defined as the causal impact of an intervention on outcomes. Treatment effect is quantitatively measured by comparing the effects of interventions on outcomes in a test group to outcomes in a control group. Randomized trials have the greatest ability to decrease bias and confounding when assessing causal effects of interventions on outcomes.

A recent Institute of Medicine (IOM) expert panel outlined the process for evaluation of biomarkers when the context of use justifies use of a biomarker as a potential surrogate endpoint [13•]. First, the performance characteristics of the test used as the biomarker should be evaluated, including standardization, and reproducibility between and within users – a process termed *analytical validity*. A test's analytical characteristics may vary under different conditions of use, such as in different patient populations or in resource-challenged settings. Therefore, analytical validity applies only to the context of use in the clinical trial setting in which the biomarker test is used.

The second step is termed *qualification*. This step consists of assessment of available evidence on relationships between the biomarker and direct outcomes for which the biomarker is a substitute.

Evidence showing correlations between changes in the biomarker and ACM independent of treatment is a necessary but not sufficient step in qualification (Fig. 1). A correlation is an association that is quantitative, such that when one variable changes, the other variable changes in a predictable and quantitative way. For example, as HIV viral load increases, ACM increases (a positive correlation) or as CD4 count increases, ACM decreases (an inverse correlation). Demonstration of correlations between the biomarker and ACM is sufficient to evaluate the biomarker as a prognostic factor, since causality is not needed for prognostic indicators. Prognostic factors select patient populations that are more or less likely to have better or worse outcomes, independent of treatment received. However, it is a different question to evaluate whether an intervention that changes a prognostic factor has a positive impact on outcomes that are causally related to the intervention. Correlations between biomarkers and ACM do not address this question, since neither associations nor correlations between biomarkers and ACM necessarily reflect *causal* effects of interventions.

Measurement of treatment effects on the biomarkers that reflect treatment effects on ACM requires causal evidence, usually from multiple randomized trials. For instance, multiple randomized clinical trials show that treatment effects of interventions that lower blood pressure reflect treatment effects of decreasing mortality, strokes, and myocardial infarctions. Multiple randomized trials show that interventions that lower viral load also lower ACM and AIDS-defining events.

Observational studies or registries that compare treated and untreated patients may have important differences in measured and unmeasured confounders at baseline between treated and untreated patients, making it challenging to assess the causality of interventions on

outcomes [14•]. Similarly, comparing subgroups of ‘responders’ to ‘non-responders’ on ACM in clinical trials and showing beneficial effects on biomarkers in the ‘responder’ group compares groups with differing baseline risks of death, making assessment of causality and outcomes difficult [15].

The utility of surrogate endpoints as a replacement for direct outcomes is contextual; that is, a surrogate endpoint that is a valid replacement for ACM under one set of conditions (e.g. types of intervention, types of patients, stage of disease, etc.) may not be valid under other conditions (e.g. interventions with different mechanisms of action). The IOM panel pointed out that the evidence for qualification of potential surrogate endpoints should be evaluated using all the evidence independent of context of use at first, and then describe conditions where the potential surrogate is useful and conditions where it is not useful [13•](p 3.18).

Application of outcome assessment principles to CM

CM is a serious and life-threatening disease; therefore, ACM is the meaningful health aspect of patients’ lives for which any biomarker in CM should serve as a replacement. Terms like ‘cure of infection’ do not explicitly state the benefit for patients in terms of how they feel, function, or survive. Irreversible morbidity such as permanent neurological defects in those who survive could also serve as a candidate for replacement by a biomarker. However, to measure neurological deficits would require developing well defined, reliable, standardized, and clinically relevant measures to evaluate this concept. To date, biomarkers in CM have not been evaluated as replacements for any other outcomes beyond ACM.

Is a biomarker needed as a surrogate endpoint for studies in CM? In the context of use in areas like sub-Saharan Africa, the mortality rate is relatively high with current therapies (one in every four to five patients treated), and death occurs in a short period of time in this acute disease. Therefore, mortality can be measured as an outcome assessment used as the primary endpoint, and the need for a biomarker as a surrogate endpoint is questionable. Recent studies show the use of ACM as an endpoint is possible [16•].

However, there has been interest, even in the context of acute disease with high short-term mortality, in using microbiological effects on fungal clearance in cerebrospinal fluid (CSF) as a surrogate endpoint for ACM in order to decrease the sample size of clinical trials or to select candidate interventions in earlier phase studies to move forward into confirmatory trials. In order to accomplish this goal, one must ascertain that a difference between test and control groups on decreasing fungal burden in CSF also shows a difference between test and control groups (treatment effects) on ACM. One must also ascertain how much of a treatment effect (effect size) on fungal burden corresponds to how much of an effect on ACM. If large treatment effects on fungal burden reflect a small benefit on ACM, then small differences in fungal burden will not reflect any meaningful difference on ACM. However, if large effects on fungal burden reflect large effects on ACM, there is little reason to use a surrogate endpoint, as ACM can be measured directly and little is gained in terms of decreasing sample size.

Cryptococcal fungi are on the causal pathway of disease in CM, therefore fungal burden in CSF is a rational biomarker to consider as a potential surrogate endpoint. However, lumbar

puncture is an invasive test and is more complex to perform, and quantitative culture is more labor intensive than biomarkers like blood testing for HIV viral load.

The biomarker outcome assessment of fungal culture in CSF has been evaluated in several ways as endpoints in CM trials; time to sterilization of CSF, proportion of patients with negative CSF cultures at specific time points, and, more recently, early fungicidal activity (EFA) [4, 17–19]. Time to sterilization of CSF has been evaluated as the time to last day of any amount of culture positivity defined as colony-forming units (CFU) of cryptococcus in CSF cultures [20]. Other trials have used the endpoint of proportion of patients with positive CSF cultures at 2 weeks post-randomization [21]. EFA is a measure of the rate of clearance of cryptococcus from the CSF, reported as a logarithm of CFU per milliliter of CSF per day. For the purpose of establishing the rate of clearance, the slope of the linear regression of log CFU/ml against time for each individual patient is determined by serial lumbar punctures, typically measured during the first 14 days after initiating antifungal therapy [22].

Although fungal cultures including EFA have been used in clinical trials, we are unaware of any studies that specifically evaluate the analytical validity of fungal cultures. Unlike obtaining blood for testing in HIV, obtaining serial lumbar punctures may have unique challenges, especially in resource-challenged settings. Furthermore, specialized equipment is necessary for growth of cryptococcus in culture, and time needed for growth and then counting of CFUs. The challenges inherent in measuring serial lumbar punctures and cultures may potentially lead to more missing data on endpoints in trials. A recent study had a number of missing data points for EFA [16•]. This may be due to challenges in obtaining the cultures or due to the competing risk of patient mortality before investigators could obtain the information.

Studies that have evaluated the relationship between EFA and ACM have focused on demonstration of correlations of patients' EFA values with mortality with various drug regimens, showing that patients who have decreasing EFA have lower ACM. These studies indicate that EFA may be useful as a prognostic indicator in that patients experiencing a slower sterilization of the CSF are likely to have a poorer outcome [18]. These studies are a useful and necessary first step in evaluation of a biomarker. However, as explained above, analyses that compare patients with various fungal outcomes are not randomized, meaning they may compare patients with inherent better prognoses based on patient, organism, or disease factors at baseline. Such analyses do not show that causal effects of treatment-induced decreases in EFA result in treatment-induced decreases in ACM, demonstrated by the difference in both EFA and ACM between the test and control group for each randomized study. These analyses would compare patients with similar baseline characteristics based on randomization, with the ability to attribute effects on EFA and ACM to the causal effects of the interventions studied. To date, individual studies in CM that demonstrate differences in EFA between the test and control groups due to treatment often show no differences between test and control groups on ACM. This may be because the individual studies are too small to detect differences in mortality or because EFA is not a valid surrogate endpoint for ACM, or that EFA is a valid surrogate only in specific contexts of use and not generally applicable.

Work is ongoing to evaluate all randomized trials that include fungal outcome assessments and ACM to perform a systematic review and meta-analysis (if possible) to compare treatment effects on EFA against treatment effects on ACM [23]. As stated by the IOM panel, the first step is to evaluate all evidence with all interventions, then follow this general analysis with analyses relevant to specific contexts of use (e.g. higher-quality studies or by types of drug regimens). One recent meta-analysis including only four studies and combined observational and randomized trials claimed a treatment effect on both EFA and ACM only at 14 days post-randomization but not at day 70 [24]. This study excluded many randomized trials in CM and does not determine whether EFA is a universally applicable surrogate endpoint for ACM. A recent randomized trial claimed a benefit on both EFA and mortality with the combination of 2 weeks of amphotericin B deoxycholate and flucytosine compared with amphotericin alone [16•]. However, this study did not adjust for multiple comparisons among multiple time points and multiple treatment regimens, reflected in the conclusions of an ‘association’ rather than a causal relationship between the interventions and outcomes [25]. Again, this single study does not show that the biomarker is universally useful as a surrogate endpoint for ACM in all contexts, and shows that the biomarker is not necessary when differences in ACM are shown. Interventions that affect host response and inflammation (e.g. steroids) may have beneficial effects on ACM even though they do not impact or adversely impact fungal clearance. This has been demonstrated in bacterial meningitis.

The possible outcomes of a systematic review/meta-analysis would be to (1) provide evidence that EFA does or does not reflect benefits on ACM, or (2) show that further studies with increased sample size are needed to confirm whether EFA is a valid surrogate endpoint for ACM if evidence is currently inconclusive. In addition, this analysis should evaluate the magnitude of the effect size on EFA and ACM, if any, as this would help to determine whether EFA can decrease the sample size of trials. If the effect size on EFA is similar to that on ACM, then there will be little benefit in decreasing sample size by using a surrogate endpoint. Investigators should also evaluate the amount of missing data, and to determine whether early mortality obviates the measurements of EFA. If this is the case, again there would be little utility to using a surrogate endpoint since the direct outcome of interest occurs first. Pre-clinical or early clinical testing may also reveal situations in CM where EFA may not be applicable as a potential surrogate endpoint, such as in interventions with a primary immunological effect rather than a direct antifungal mechanism.

The process of evaluating a potential surrogate endpoint includes specifying the context of use to determine if that context warrants a surrogate endpoint (e.g. chronic disease such as HIV/AIDS, where outcomes like ACM occur years in the future, or acute disease where direct outcomes are rare). Another aspect of context of use would be the mechanism of action of the intervention. Interventions whose mechanism of action is primarily immunological may not have an effect on microbiological endpoints that reflects effects on direct outcomes like ACM. Investigators should clearly define the aspect of patient’s life (survival or irreversible morbidity) for which the surrogate is a substitute. The potential surrogate should be on the causal pathway of the disease. Potential surrogates should also show a correlation between changes in the potential surrogate and the direct outcome of interest (ACM) as a necessary but not sufficient first step, since this correlation may not be

causal and may compare sicker with less sick patients. Finally, the most important step in evaluating a potential surrogate is the comparison of treatment effects (the difference in outcomes between the test and control groups) in randomized clinical trials on both the potential surrogate and the direct outcome of interest (ACM) to show that the potential surrogate can be used as a substitute for effects of the intervention on the direct outcome. Many surrogates show correlations with direct outcomes that, on further evaluation, are not causally related to treatment effects of interventions.

The IOM panel pointed out “strong evidence and a compelling context are needed for the utilization of a biomarker as a surrogate endpoint.” [13•](page S-8). At the present time, it appears work is incomplete and still ongoing to qualify fungal outcomes, including EFA, as a valid replacement for ACM in CM. For the time being, studies should be based on the primary outcome of ACM, with sufficient sample sizes and focused research questions with clear and quantifiable measurements to detect differences of various regimens on ACM. Using appropriate methodology to measure both direct effects of interventions on ACM as well as properly developing, evaluating, and qualifying potential surrogate endpoints in the appropriate context will help clinicians choose the best treatment options for patients with CM.

Acknowledgments

This project has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health, under contract N01-CO-12400. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government.

This research was supported in part by the National Institute of Allergy and Infectious Disease.

References and Recommended Reading

Papers of particular interest, published recently, have been highlighted as:

- Of importance

1. Park BJ, Wannemuehler KA, Marston BJ, Govender N, Pappas PG, Chiller TM. Estimation of the current global burden of cryptococcal meningitis among persons living with HIV/AIDS. *AIDS*. 2009; 23(4):525–30. [PubMed: 19182676]
2. Harrison TS. The burden of HIV-associated cryptococcal disease. *AIDS*. 2009; 23(4):531–2. [PubMed: 19240459]
3. Kambugu A, Meya DB, Rhein J, O'Brien M, Janoff EN, Ronald AR, et al. Outcomes of cryptococcal meningitis in Uganda before and after the availability of highly active antiretroviral therapy. *Clin Infect Dis*. 2008; 46(11):1694–701. [PubMed: 18433339]
4. Day JN, Chau TT, Lalloo DG. Combination antifungal therapy for cryptococcal meningitis. *N Engl J Med*. 2013; 368(26):2522–3. [PubMed: 23802521]
5. Kirch W, Schaffii C. Misdiagnosis at a university hospital in 4 medical eras. *Medicine (Baltimore)*. 1996; 75(1):29–40. [PubMed: 8569468]
6. Altman DG, Bland JM. How to randomise. *BMJ*. 1999; 319(7211):703–4. [PubMed: 10480833]
7. Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? *Ann Intern Med*. 1996; 125(7):605–13. [PubMed: 8815760]
- 8•. Fleming TR, Powers JH. Biomarkers and surrogate endpoints in clinical trials. *Stat Med*. 2012; 31(25):2973–84. Explanation of the issues associated with the choice of biomarkers as surrogate

endpoints, the pros and cons of the use of surrogates, and examples of successful and failed biomarkers as surrogate endpoints in randomized trials. [PubMed: 22711298]

9. Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther.* 2001; 69(3):89–95. [PubMed: 11240971]
10. Walton MK, Powers JH, Hobart J, Patrick DL, Marquis P, Vamvakas S, et al. Clinical outcome assessments: conceptual foundations. *Value in Health.* 2013 in press.
11. Food US, Administration D. Code of federal regulations. Subpart H – accelerated approval of new drugs for serious and life-threatening illnesses. Silver Spring: US FDA; 1992.
12. Avorn J. Approval of a tuberculosis drug based on a paradoxical surrogate measure. *JAMA.* 2013; 309(13):1349–50. [PubMed: 23430122]
13. Institute of Medicine. Evaluation of biomarkers and surrogate endpoints in chronic disease. Washington, DC: National Academies Press; 2010. Long but thorough discussion about the development process for potential surrogate endpoints
14. Hartz A, He T, Wallace R, Powers J. Comparing hormone therapy effects in two RCTs and two large observational studies that used similar methods for comprehensive data collection and outcome assessment. *BMJ Open.* 2013; 3(7) Example of lack of ability to control for unmeasured confounders in observational studies despite controlling for over 900 baseline variables.
15. Yusuf S, Wittes J, Probstfield J, Tyroler HA. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. *JAMA.* 1991; 266(1):93–8. [PubMed: 2046134]
16. Day JN, Chau TT, Wolbers M, Mai PP, Dung NT, Mai NH, et al. Combination antifungal therapy for cryptococcal meningitis. *N Engl J Med.* 2013; 368(14):1291–302. Clinical trial using all-cause mortality as an endpoint in cryptococcal meningitis. [PubMed: 23550668]
17. Bicanic T, Meintjes G, Wood R, Hayes M, Rebe K, Bekker LG, et al. Fungal burden, early fungicidal activity, and outcome in cryptococcal meningitis in antiretroviral-naïve or antiretroviral-experienced patients treated with amphotericin B or fluconazole. *Clin Infect Dis.* 2007; 45(1):76–80. [PubMed: 17554704]
18. Bicanic T, Muzoora C, Brouwer AE, Meintjes G, Longley N, Taseera K, et al. Independent association between rate of clearance of infection and clinical outcome of HIV-associated cryptococcal meningitis: analysis of a combined cohort of 262 patients. *Clin Infect Dis.* 2009; 49(5):702–9. [PubMed: 19613840]
19. Jarvis JN, Meintjes G, Rebe K, Williams GN, Bicanic T, Williams A, et al. Adjunctive interferon-gamma immunotherapy for the treatment of HIV-associated cryptococcal meningitis: a randomized controlled trial. *AIDS.* 2012; 26(9):1105–13. [PubMed: 22421244]
20. Bennett JE, Dismukes WE, Duma RJ, Medoff G, Sande MA, Gallis H, et al. A comparison of amphotericin B alone and combined with flucytosine in the treatment of cryptococcal meningitis. *N Engl J Med.* 1979; 301(3):126–31. [PubMed: 449951]
21. van der Horst CM, Saag MS, Cloud GA, Hamill RJ, Graybill JR, Sobel JD, et al. Treatment of cryptococcal meningitis associated with the acquired immunodeficiency syndrome. National Institute of Allergy and Infectious Diseases Mycoses Study Group and AIDS Clinical Trials Group. *N Engl J Med.* 1997; 337(1):15–21. [PubMed: 9203426]
22. Brouwer AE, Rajanuwong A, Chierakul W, Griffin GE, Larsen RA, White NJ, et al. Combination antifungal therapies for HIV-associated cryptococcal meningitis: a randomised trial. *Lancet.* 2004; 363(9423):1764–7. [PubMed: 15172774]
23. Montezuma-Rusca, M.; Powers, JH.; Williamson, P.; Sullivan, B.; DerSimonian, R. PROSPERO registry of systematic reviews. University of York; 2014. Evaluation of fungal clearance as a surrogate endpoint for mortality in the treatment of cryptococcal meningitis.
24. Yao ZW, Lu X, Shen C, Lin DF. Comparison of flucytosine and fluconazole combined with amphotericin B for the treatment of HIV-associated cryptococcal meningitis: a systematic review and meta-analysis. *Eur J Clin Microbiol Infect Dis.* 2014; 33(8):1339–40. [PubMed: 24550039]
25. Proschan MA, Waclawiw MA. Practical guidelines for multiplicity adjustment in clinical trials. *Control Clin Trials.* 2000; 21(6):527–39. [PubMed: 11146147]

Figure 1.

Panel (a) shows how randomization at the baseline of a clinical trial equally distributes patients with good prognosis (*green*) and poor prognosis (*red*) between groups. Panel (b) shows that, when followed to outcomes, randomization allows assessment of causal effects (white dot = successful outcome, no dot = treatment failure) so that any effects demonstrated are due to the interventions tested and not to confounding based on baseline prognostic factors, since these remain balanced between the groups. Panel (c) shows the common method of evaluating correlations for prognostic factors, in which patients with ‘successful’ outcomes are compared with patients with ‘unsuccessful’ outcomes. Note these analyses are not randomized and compare patients with inherently better prognosis (*green*) with patients with inherently worse prognosis (*red*), meaning that the differences measured may not be due to the interventions tested but to baseline differences between patients. Panel (d) shows the appropriate way to evaluate biomarkers, where the treatment difference between randomized test and control groups are compared for effects on the biomarker outcome (e.g. early fungicidal activity) and effects on the direct outcome like all-cause mortality. In this example, there is a large effect (42 %) on the biomarker, but no effect on the direct outcome of interest (0 %), showing that this biomarker is not valid for use as a surrogate endpoint.