Published in final edited form as:

Stat Commun Infect Dis. 2017 ; 9(1): . doi:10.1515/scid-2016-0003.

Evaluating Anti-Infective Drugs in the Resistant Pathogen Setting: Can we Use External Controls?

Scott R Evans¹ and John Powers^{2,3}

¹Center for Biostatistics in AIDS Research and the Department of Biostatistics, Harvard University, USA

²Leidos Biomedical Research in support of the Division of Clinical Research, National Institutes of Health, Bethesda, Maryland, USA

³George Washington University School of Medicine, Washington D.C., USA

Abstract

Decreased efficacy of antibiotics due to resistant pathogens has created a need for the development of more effective medical interventions. Despite the increasing prevalence of pathogens resistant to one or more drugs, identifying and enrolling participants into clinical trials that evaluate new interventions for the treatment of some diseases can be challenging given the low prevalence of disease in which there are no effective treatments. Thus researchers might be tempted to consider externally-controlled trials that may allow for a reduction of the necessary number of prospectively-identified trial participants, thus easing recruitment burden and resulting in more timely trial completion relative to randomized controlled trials. We discuss advantages and disadvantages in externally controlled trials and review requirements for a valid externally-controlled trial. As ECTs are subject to the bias of observational studies, the criteria for a valid ECT should be carefully evaluated before these designs are implemented. Given considerable variation in study results in the resistant pathogen setting, the lack of information on important patient characteristics that may confound estimates of treatment effects, as well as the improvements in medical practice and evolving antibiotic resistance, the use of ECTs in the resistant pathogen setting, is not recommended. ECTs should be limited to specific situations where superiority of the effect of the new intervention is dramatic, the usual course of the disease highly predictable, the endpoints are objective (e.g., all-cause mortality) and the impact of baseline and treatment variables on outcomes is well characterized. Given that the resistant pathogen setting does not satisfy these criteria, we conclude that randomized clinical trials are needed to evaluate new treatments for resistant pathogens. Innovative approaches to trial design that may ease recruitment burden while evaluating the benefits and harms of new treatments are being developed and utilized.

Corresponding author contact information: Scott R Evans, PhD, FXB-625A, Center for Biostatistics in AIDS Research and the Department of Biostatistics, Harvard TC Chan School of Public Health, 651 Huntington Avenue, Boston, Massachusetts, USA, PH: 617-432-2998, FAX: 617-432-3163, evans@sdac.harvard.edu.

Conflicts of Interest: None.

Keywords

Historical controls; external controls; randomization; antibiotics; drug resistant pathogens

Introduction

Decreased efficacy of antibiotics on reducing mortality and morbidity due to resistant pathogens has created a need for the development of more effective medical interventions. However despite the increasing prevalence of resistant pathogens, identifying and enrolling participants into clinical trials that evaluate new interventions for the treatment of some diseases can be challenging given the low prevalence of disease in which there are no effective options. Thus researchers might consider alternative designs that require fewer trial participants and can produce results more quickly but still maintain scientific validity. One design option for consideration is the use of externally-controlled trials (ECTs). In some situations, ECTs may allow for a reduction of the necessary number of prospectively-identified trial participants, thus easing recruitment burden and resulting in more timely trial completion relative to randomized controlled trials (RCTs). Can we use external controls to obtain valid evidence in the resistant pathogen setting?

Definitions

ECTs are non-randomized designs that can take one of two forms: (1) historical-controlled trials (HCTs), or (2) concurrent external controlled trials (CECTs). The difference between HCTs and CECTs is the time period in which control group data were collected – either retrospectively (for HCTs), e.g., from completed studies published in the medical literature or reviews of patient charts, or concurrently (CECTs) with the prospective data collection for test intervention arm participants. CECTs may provide information that is more relevant to current practice. During the analysis of ECTs, the external control is compared with the prospective test arm with respect to important endpoints of interest.

Advantages and Disadvantages of ECTs

An advantage of using ECTs is that the trial may require fewer prospectively enrolled participants due to the absence of a randomized concurrent control group and thus provides resource and time efficiency. However, if the hypothesized size of the treatment effect is conservative or more stringent criteria for statistical significance are used then there may be little difference between RCTs and ECTs with regard to the numbers of patients enrolled. ECTs are often considered in trials evaluating therapies to treat rare diseases when the patient pool is limited or when randomization to the control intervention is undesirable in the absence of effective treatments. In some situations it may also be easier to recruit potential trial participants for the prospective component of an ECT as participants may prefer knowing which intervention they will receive. In RCTs, participants do not have a choice regarding which intervention they receive. In RCTs, even after the intervention is assigned; participants may remain blinded to intervention assignment. Conversely participants enrolled prospectively into ECTs will be aware of their treatment assignment potentially making participation more attractive (Table 1).

The major drawback of ECTs is that they are non-randomized studies. Randomization is the foundation for statistical inference, providing the theoretical foundation for valid estimation of treatment effects and the testing of hypotheses. Randomization ensures the expectation of between-arm balance with respect to all factors, known or unknown, measured or unmeasured. Without randomization, estimates of treatment effects can be seriously biased and tests of treatment effects invalid. Because they are non-randomized studies, ECTs are potentially vulnerable to all of the biases of observational studies.

Bias can occur if the controls systematically differ from the prospective test group with regard to important factors in a manner that can affect outcome. Differences can occur due to participant selection (e.g., patients with more favorable prognosis or lesser disease severity are selected for the prospective component of the trial but poor risk patients are not enthusiastically recruited or are excluded due to knowledge of the outcome), supportive care, concomitant therapies, follow-up strategies, and evaluation methods. It can also be challenging to define a “time zero” representing a baseline in the external component of the ECT, to align with time zero in the prospective component of the study. If the ECT is an HCT, then bias may occur due to factors that have changed since the time the historical control group data were collected (e.g., improvement in medical practice and patient standard of care, diagnostic criteria, or referral patterns). Furthermore ECTs are not blinded and thus are subject to bias when eligibility for enrollment, inclusion in analyses, or outcomes are assessed by clinicians or patients. In ECTs, clinicians may also selectively prescribe additional therapies given the knowledge of the treatment assignment. ECTs utilizing “control groups” that are not well-defined and based on “general knowledge” are particularly problematic given the inability to compare patient characteristics between the test intervention group and external control group. Researchers should try to identify participants for the ECT control group who otherwise would meet entry criteria for an RCT.

ECTs have been used rarely in clinical trials for late-stage drug development due to the concerns for these potential biases. International guidance (ICH-E10) recommends reserving ECTs for specific situations where the effect of interventions is large and highly unlikely to be explained by other factors. For example in the pre-antibiotic era, bacterial endocarditis had a 100% mortality rate (but prior to early diagnosis by echocardiogram and other changes in supportive care and surgery). A large reduction in mortality in an ECT compared to a current external control group might be convincing. ICH-E10 also recommends that the usual course of the disease is highly predictable and independent of patient factors. It is further recommended that ECTs be limited to cases in which the endpoints are objective (e.g. all-cause mortality, or relapse of disease if rigorously defined on objective criteria) to avoid subjective evaluations given the unblinded nature of ECT trials. During consideration of conducting an ECT, researchers should carefully evaluate the requirements for a valid ECT (Table 2).

The Validity of ECTs

Research suggests that ECTs tend to produce “positive” results more frequently than RCTs. Sacks et. al. (1983) reviewed 50 RCTs and 56 HCTs evaluating the same six therapies and found 79% of the HCTs but only 20% of the RCTs demonstrated superiority of the test

group to the control. These differences were largely due to the worse outcomes in the control groups in HCTs compared to the corresponding control groups in RCTs. This is not surprising since RCTs typically have more select entry criteria than the control groups in HCTs (i.e., RCTs typically exclude seriously ill patients but such patients may be included in the control groups of HCTs).

The infectious disease community may be familiar with the story of patulin, a metabolic product of *Penicillium patulum* Bainier which illustrates the issues with ECTs. Patulin was studied for the treatment of the common cold in a non-randomized, concurrent externally controlled (patulin in buffer vs. buffer alone) clinical trial of 180 subjects in 1943. The number of subjects that improved on subjective non-standardized assessments of symptoms at 48 hours in the patulin in buffer arm was 55/95 (58%) vs. 8/85 (9.4%) in buffer alone arm, representing a difference of 48%, 95% CI=(35%, 60%), $p<0.002$. The dramatic results triggered a follow-up study, a randomized, controlled, double-blind trial in 1449 factory and postal workers, one of the first randomized trials ever. The results of this trial were quite different. The number of subjects cured at 48 hours in the patulin in buffer arm was 87/668 (13%) vs. 88/680 (13%) in the buffer alone arm, resulting in a difference of 0%, 95% CI=(-3.6%, 3.8%), $p = 0.96$ (Patulin Clinical Trials Committee, 2004; Lancet 1943).

Other examples of HCTs and RCTs having reached differing conclusions include the Women's Health Initiative in which RCTs revealed that post-menopausal hormone supplementation provided an adverse effect on cardiovascular outcomes in contrast to HCTs that found beneficial effects on cardiovascular outcomes (Writing Group for the Women's Health Initiative Investigators, 2002). Nobel Laureate Linus Pauling published results of HCTs indicating that high-dose vitamin C lengthened survival among pre-terminal cancer patients (Cameron and Pauling, 1976; Cameron and Pauling 1978). However a RCT concluded no such benefit (Moertel et. al., 1985).

The validity of an ECT depends on the assumption that controls have the same distribution of important baseline characteristics compared to the participants in the test intervention arm. Appropriate analyses include a between-group comparison of these characteristics that can potentially confound the results if imbalanced. Frequently the assumption of a balance of important baseline factors does not hold or is difficult to confirm.

Statistical methods that adjust for the potentially confounding effects of imbalances are often utilized. For example, propensity scores may be used to adjust for differences in patient characteristics based on baseline characteristics that are known and measured. Study participants are stratified by their propensity score, investigators compare patients within a propensity score, and results are pooled across propensity categories. Multivariable regression modeling, stratification, restriction, matching, or instrumental variable methods may also be utilized (Stafford et. al, 2014). Most of these adjustments require raw patient-level data, which are sometimes unavailable for the controls. Unfortunately using modern statistical methodologies cannot completely address the biases associated with ECTs and invariably depend on untestable assumptions. Some important factors may be beyond our current medical understanding and thus unknown or may not have been measured in the ECT (i.e., known and measured variables are only the tip of the iceberg) resulting in

unmeasured confounding and making it impossible to adjust for these factors. For example, a recent study evaluated the differences in outcomes between RCTs and observational studies studying the effects of hormone replacement therapy. Despite adjusting for more than 900 baseline variables, differences in outcomes between the two designs were unable to be explained (Hartz et. al., 2013).

Can we use ECTs in the resistant-pathogen setting?

The requirements for a valid ECT are challenging to demonstrate in the resistant-pathogen setting. For many infectious diseases even in the setting of susceptible pathogens, patient factors (e.g., age, severity of illness) play an important role in explaining observed outcomes independent of the infecting pathogen. Unfortunately many important variables were not measured in historical studies, making it impossible to control for these factors if this information was used in a future ECT.

In addition, medical practice is constantly changing. In a 10-year longitudinal study conducted at a single ICU (Rosenburger et. al., 2012), the mortality rate decreased despite the rise of resistant bacterial infections. The authors attributed the decrease in mortality to improvements in technology and critical care. Supportive care other than antibiotics may also impact outcomes and differ geographically as noted recently in Ebola Virus Disease (Cox et. al., 2014). Such improvements in standard medical practice would confound the results of HCTs such that observed decreases in mortality could not be attributed to the interventions being tested.

Resistance patterns are also constantly evolving. Trials conducted in the future will be conducted in diseases caused by organisms with different resistance profiles and different patient characteristics compared to past trials. This again would violate the requirement that controls and test group participants should have the same distributions of important baseline factors. Of note the temporal and geographic variation of disease also makes this disease area unsuitable for response adaptive randomization. Revision of the randomization ratio could threaten the expectation of balance of temporal and geographic factors originally ensured by randomization.

As with the case with pneumococcal pneumonia in the 1940's, current evidence in the resistant pathogen setting indicates that historical studies display inconsistent results on objective outcomes such as all-cause mortality, challenging the validity of an HCT. Researchers may assume that in the setting of resistance to all available antibiotics, that the mortality rate will be consistently high. However, a meta-analysis of deaths attributable to carbapenem-resistant *Enterobacteriaceae* infections (Falagas et al. 2014) showed non-uniform failure and substantial variation in mortality outcomes, with point estimates for survival ranging from 6% to 70% across nine studies. Given the clinical heterogeneity and variation of mortality outcomes in these settings, the interpretation of the results of future HCTs would be very difficult to put into context and interpret.

Can RCTs be Done?

Given the inadequacies of ECTs for evaluating interventions for resistant infections, it is important to focus on developing innovative approaches to conducting RCTs.

Many claim that RCTs will be unable to enroll enough trial participants as some studies have struggled with slow accrual rates. However this appears to be in conflict with the observation that resistant pathogens are becoming increasingly common. It is also difficult to understand why an RCT would not be able to be conducted but a study using concurrent external controls would be feasible.

One challenge to completing enrollment in trials evaluating treatments for resistant pathogens has been on the use of noninferiority (NI) clinical trial designs. NI designs often have strict entry criteria (e.g., limitations on prior therapy) to preserve assay sensitivity, a requirement for valid NI trials, and may require large sample sizes. If superiority trials could be conducted then less strict entry criteria could be implemented and sample sizes may be reduced.

Despite the challenges of enrollment faced by prior RCTs, RCTs can be conducted in the resistant pathogen setting. Innovative statistical approaches as well as improved design and study planning can help to accomplish this goal.

Innovative statistical methodologies (Evans and Follmann, 2015; Evans and Follmann, 2016) that evaluate the effects of interventions for the treatment of infections due to resistant pathogens are being developed and applied. The methodologies may reduce the required sample sizes and utilize superiority designs, allowing for more liberal entry criteria (e.g., reduced prior therapy restrictions). For example, Evans and Follmann (2016) introduce the “partial credit” design and discuss an example of a trial evaluating a new intervention to treat resistant infections. The new intervention was hypothesized to have efficacy advantages as well as reduced toxicity. Use of a composite benefit:risk outcome comprised of efficacy and toxicity potentially allows for reduced sample size and synthesized evaluation of benefits and harms with a superiority design. The Antibacterial Resistance Leadership Group (ARLG) is designing a trial using the partial credit strategy.

Future work might evaluate whether data from RCTs can be augmented by external data in selected cases for resistant pathogens. For example, in sites expecting to enroll in a future RCT, observational data might be collected to gain knowledge of temporal and geographic variation of disease and patient characteristics that affect outcomes. These data may inform control for confounding when augmenting future RCT data with external data from the same sites utilized in the RCT.

A site recruitment strategy could utilize high-performing sites from recent studies and leverage existing networks such as the Antibacterial Resistance Leadership Group (ARLG) and the Combatting Bacterial Resistance in Europe - Carbapenem Resistance (COMBACTE-CARE). Emerging rapid diagnostics may help target high-risk patients.

Conclusions

ECTs are subject to the bias of observational studies. The criteria for a valid ECT should be carefully evaluated before these designs are chosen and implemented. Misleading information resulting from misguided ECTs could be worse than having limited data since patients could be harmed as a result of erroneous conclusions. Theoretically an ECT could be valid in a resistant-pathogen setting if outcomes were uniform on an objective endpoint such as all-cause mortality and independent of patient factors, and if the control group received the best current standard of care and was matched closely for enrollment criteria and other supportive care to evaluate the superiority of a new intervention. Noninferiority hypotheses would not be appropriate since in the setting of noninferiority, patients already have an effective intervention (the control drug) and the biases of an ECT would not be ethically or scientifically acceptable.

However we are not aware of scenarios in the resistant pathogen setting where these criteria for valid ECTs are met or are likely to be arise in the near future. Given the considerable variation in study results in the resistant pathogen setting, the lack of information on important patient characteristics that may confound estimates of treatment effects, as well as the improvements in medical practice and evolving antibiotic resistance, the use of ECTs in the resistant pathogen setting is not recommended.

Given the issues with non-randomized evidence, it is important to focus on developing innovative approaches to RCTs. Methodologies such as “partial credit” and use of superiority designs are being developed and employed.

Acknowledgments

Research reported in this publication was supported by the National Institute of Allergy And Infectious Diseases of the National Institutes of Health under Award Number UM1AI104681. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

- Cameron E, Pauling L. Supplemental ascorbate in the supportive treatment of cancer: prolongation of survival times in terminal human cancer. *Proc Natl Acad Sci USA*. 1976; 73:3685–3689. [PubMed: 1068480]
- Cameron E, Pauling L. Supplemental ascorbate in the supportive treatment of cancer: reevaluation of prolongation of survival times in terminal human cancer. *Proc Natl Acad Sci USA*. 1978; 75:4538–4542. [PubMed: 279931]
- Cox E, Borio K, Temple R. Evaluating Ebola Therapies – the Case for RCTs. *N Engl J Med*. 2014; 371(25):2350–1. [PubMed: 25470568]
- Evans SR, Follmann D. Using Outcomes to Analyze Patients Rather than Patients to Analyze Outcomes: A Step toward Pragmatism in Benefit:risk Evaluation. *Statistics in Biopharmaceutical Research*. 2016
- Evans SR, Follmann D. Fundamentals and Innovation in Antibiotic Trials. *Statistics in Biopharmaceutical Research*. 2015; 7(4):331–336. DOI: 10.1080/19466315.2015.1094406 [PubMed: 27087893]
- Falagas ME, Tansarli GS, Karageorgopoulos DE, Vardakas KZ. Deaths attributable to carbapenem-resistant *Enterobacteriaceae* infections. *Emerging Infectious Diseases*. 2014; 20(7)doi: 10.3201/eid2007.121004

- Gehan EA. The evaluation of therapies: historical control studies. *Statistics in Medicine*. 1984; 3:315–324. [PubMed: 6528131]
- Hartz A, He T, Wallace R, et al. 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:e002556.doi: 10.1136/bmjopen-2013-002556
- International Conference on Harmonization (ICH)-E10. Choice of Control Group and Related Issues in Clinical trials. 2000
- Moertel CC, Fleming TR, Creagan ET, et al. High dose vitamin C and placebo in the treatment of advanced cancer patients with no prior chemotherapy. A randomized double-blind comparison. *NEJM*. 1985; 312:137–141. [PubMed: 3880867]
- Patulin Clinical Trials Committee, Medical Research Council. Clinical trial of patulin in the common cold. *International Journal of Epidemiology*. 2004; 33:243–246. DOI: 10.1093/ije/dyh028 [PubMed: 15082620]
- Patulin in the Common Cold: Introduction by H. Raistrick; Biochemistry and Chemistry, by JH Birkinshaw, SE Michael, A Bracken, and H Raistrick; Preliminary Trial in the Common Cold, by WE Gye; Biological Properties and Extended Trial in the Common Cold, by WA Hopkins; Statistical Note by M Greenwood. *Lancet*. 1943; ii:625.
- Rosenberger LH, Lapper DJ, Sawyer RG. Infections caused by multidrug resistant organisms are not associated with overall, all-cause mortality in the surgical intensive care unit: the 20,000 foot view. *J Am Coll Surg*. 2012; 214:747–755. [PubMed: 22421258]
- Sacks H, Chalmers TC, Smith H. Randomized versus historical controls for clinical trials. *Am J Med*. 1982; 72:233–240. [PubMed: 7058834]
- Stafford KA, Boutin M, Evans SR, Harris AD. Difficulties in demonstrating superiority of an antibiotic for multi-drug resistant bacteria in non-randomized studies. *CID*. 2014; 59(8):1142–7.
- Writing Group for the Women’s Health Initiative Investigators. Risks and benefits of estrogen plus progestin in healthy menopausal women: principle results from the Women’s Health Initiative Randomized Controlled Trial. *JAMA*. 2002; 288:321–333. [PubMed: 12117397]

Table 1**Advantages and Disadvantages of Externally-Controlled Trials**

Advantages	
<ul style="list-style-type: none">• Fewer prospectively enrolled participants• Faster trial results• Less expensive• Potentially more attractive for trial participants given knowledge of treatment assignment	
Disadvantages	
<ul style="list-style-type: none">• Non-randomized studies, thus subject to the biases of observational studies• Bias can occur if the external controls systematically differ from the prospective group in a manner that can affect outcome (e.g., differences due to participant selection (e.g., patients with more favorable prognosis or disease severity are selected for the prospective trial but poor risk patients are not enthusiastically recruited; differing demographics), supporting care, concomitant therapies, follow-up, and evaluation methods)• If the external controls are historical instead of concurrent, then bias may occur due to factors that have changed since the time that historical study was conducted (e.g., improvement in medical practice, diagnostic criteria, or referral patterns)• Not blinded and thus are subject to selective entry as well as patient and evaluator bias when assessing eligibility and outcomes• Biases cannot be completely addressed using modern statistical methodologies	

Table 2

Requirements for a Valid Externally-Controlled Trial (adapted from Gehan, 1984 and ICH-E10)

- The control group has received a well-defined treatment
- The control group was either: (1) concurrent, or (2) historical but disease and treatment conditions have not changed since the historical study was conducted
- The criteria for eligibility, observation, and evaluation methodology are the same for both groups
- The external study collected high-quality and reliable data
- Patient-level data on known confounding variables are available for analyses
- There are no unknown confounders
- The statistical model used to control for confounding is correctly specified
- External studies demonstrate consistent and reproducible results
- If between-group differences in prognostic variables exist, then they are not of sufficient magnitude to explain observed differences in outcome
- Endpoints are objective (e.g., all-cause mortality)