A NEW TOOLKIT FOR CONDUCTING CLINICAL TRIALS IN RARE DISORDERS

Lusine Abrahamyan[1], Ivan R Diamond[2], Sindhu R Johnson[3], Brian M Feldman[4]

[1]Institute for Clinical Evaluative Sciences, University of Toronto, Toronto, ON, Canada; [2]The Division of General Surgery, University of Toronto, Toronto, ON, Canada; [3]Department of Medicine, Toronto Western and Mount Sinai Hospitals, University of Toronto, Toronto, ON, Canada; [4]Department of Pediatrics, The Hospital for Sick Children, University of Toronto, Toronto, ON, Canada

Corresponding Author: *brian.feldman@sickkids.ca*

ABSTRACT

Evidence based medicine requires strong scientific evidence upon which to base treatment. Because the available study populations for rare diseases are small, this evidence is difficult to accrue. Investigators need to consider a flexible toolkit of methods to deal with the problems inherent in the study of rare disease. We present some potential solutions in this paper.

Key Words: *Evidence-based medicine, observational studies, rare disorders, NNT, adaptive trial design, Bayesian analysis*

INTRODUCTION

Rare diseases are estimated to affect millions of North Americans and millions globally.[1] Although any one rare disease may affect only few patients, in aggregate the public health impact is huge. The Institute of Medicine has identified a number of challenges in developing safe and effective treatments for rare diseases, including difficulties in attracting research funding, recruiting sufficient numbers of research subjects, finding appropriate research designs for small populations, and securing adequate expertise at regulatory agencies for the approval of new products. [1]

Reasoned clinical decision making depends on having accurate information. Rational therapeutic decisions weigh the benefits and costs (in terms of adverse outcomes as well as financial costs) of the alternatives; better decisions can be made with valid information about the benefits and costs of alternative treatments.

Modern clinical science is an inductive practice. The inductive method was proposed by Sir Frances Bacon – to replace Aristotle's deductive syllogism. In the inductive method, we make observations about nature, use them to develop hypotheses, and then test those hypotheses. A strong inductive argument is one in which the evidence supplied by its premises makes it highly improbable that its conclusion is false when all premises are true.[2]

The inductive method was applied to clinical medicine by Parisian physician, Pierre Louis, in the early 1800s. He devised his numerical method, and used it to show that bloodletting (the application of leeches) led to increased mortality in pneumonitis.[3] His influential publication led to the disappearance of leeching worldwide. This was one of the first applications of evidence-based practice.

Evidence-based medicine (EBM), while an old practice, is a relatively new term coined in 1990.[4] It is defined as practicing medicine using expert clinical judgment, combined with the best external evidence, and guided by patient values.[5] Evidence that is derived from observational data can suffer from confounding, resulting in a distortion of the estimated treatment effect.[6] The confounder is both a) causally related to the outcome independent of the exposure, and b) associated with the exposure but not a consequence of exposure.

One important source of bias in observational treatment studies is *confounding by indication* (also known as treatment selection bias or susceptibility bias).[6,7] In this situation prognostic factors actually influence treatment

exposure, i.e., patients with a better prognosis receive one treatment while patients with a worse prognosis receive another treatment.[7]

However, some of the bias is reduced by statistical adjustments if the confounding factors are both known and measured.[6]

Clinical experiments (e.g. randomized controlled trials – Fig. 1) are considered to be the most valid way to generate unbiased evidence, and avoid confounding. Imbalances between groups are reduced to chance, and this probability can be made small by increasing the sample size.

However, in rare diseases a large sample size for study is not readily available. Thus, observational studies may have special importance for rare diseases. Observational studies may also evaluate treatment efficacy in a population that is more representative than persons in a randomized clinical trial.[8]

There are an estimated 5,000 to 8,000 rare or orphan diseases that affect about 25 million Americans and 30 million Europeans.[9] Although there is no single definition of rare diseases, countries and organizations define it by their prevalence. For example, in the European Union a disease is called rare if it affects no more than 5 out of 10,000 in the population.[9] In the US, a rare disease affects 7 out of every 10,000.[10] Many or most of the rare diseases have a genetic origin, start their manifestation early in childhood, are severe, chronic and often life-threatening with high psychosocial burden and poor quality of life.[11]

We have little evidence for the treatment of rare diseases due to several inherent constraints to internal and external validity of small (low sample size) trials.[12] Research standards and clinical trial protocol requirements are the same for both rare and common diseases; however, because of a low incidence and prevalence, patient accrual into RCTs of rare diseases may be very difficult and almost infeasible in some instances. For example, an RCT involving treatment of indolent T-cell LGL leukemia would need 439 patients to demonstrate a 50% relative risk reduction of death with a total trial duration of 5 years. However, considering the usual 5% patient enrolment into cancer trials, and a disease prevalence of 160 patients per year, it would take up to 55 years to enroll the required number of patients.[13] The RCT of itraconazole for the prevention of severe fungal infection in children and adults with chronic granulomatous disease required 10 years to enroll the required sample of 39 subjects.[14]

Because rare diseases may have variable disease progressions, it may be hard to enroll a homogeneous study population and achieve balanced randomization. Selecting a more heterogeneous population may increase the study's external validity but may also lead to erroneous conclusions. For example, when patients with different sub-types of peripheral T-cell lymphoma were considered together, investigators achieved inaccurate conclusions about the disease prognosis.[13]

Because the disease progression in many rare diseases is not fully described, investigators may need to use composite or surrogate outcomes with un-established validity and reliability. For example, for many inborn errors of metabolism the levels of various biomarkers as predictors of good outcome are not well established.[15]

The quality of RCTs in rare diseases may be jeopardized because of the existing constraints, as described above. For example, a systematic review of RCTs in juvenile idiopathic arthritis (JIA) found that only about 5% of studies met the 6 included quality indicators.[16]

Design studies to only pick up large treatment effects

Perhaps one solution to the study of treatment for rare diseases is to only design trials that will uncover very large treatment effects (far fewer subjects are needed in RCTs to demonstrate large treatment effects). For common diseases, treatments of only modest effect may not benefit many individual patients, but may offer a large benefit to society. For rare diseases, we may only be interested in treatments that are likely to benefit our individual patients themselves.

The number needed to treat (NNT) can be used to illustrate this point. The NNT is the number of patients that a physician needs to treat in order to affect one additional cure, or response.[17] For illnesses where the treatment benefit is obtained by a few (e.g., NNT = 100, 1

additional patient in 100 benefits), if the treatment was relatively affordable and safe enough, it would be wise public policy to treat everyone – even though the chance that any one individual would benefit from treatment is low. For a very rare disease, and for a treatment with a similar NNT, there would be no public policy argument for treatment, and since most patients would not see a personal benefit, the wise choice may be to not offer treatment. For very rare diseases, we might argue that we are only interested in treatments that are highly likely to benefit the individual patient – perhaps with NNTs in the order of 2 or 3. This is likely a controversial solution, and warrants further discussion and debate.

Use more acceptable study designs (i.e. limited exposure to placebo)

Study designs that have higher acceptability among investigators and patients may have better enrolment, an important factor for studies of rare diseases. Acceptability can be improved by eliminating a placebo comparison or by allowing it only for a short time. Here we present some examples of such designs.

Active comparator

Patients may not wish to enroll into trials because of concerns about being randomized to a placebo.[18] The most common types of control conditions used in RCTs are placebo controls; active comparator controls are used when there is a currently available acceptably effective and safe treatment. A survey that evaluated physicians' preferences showed that physicians were significantly more likely to enroll into active-controlled trials.[19]

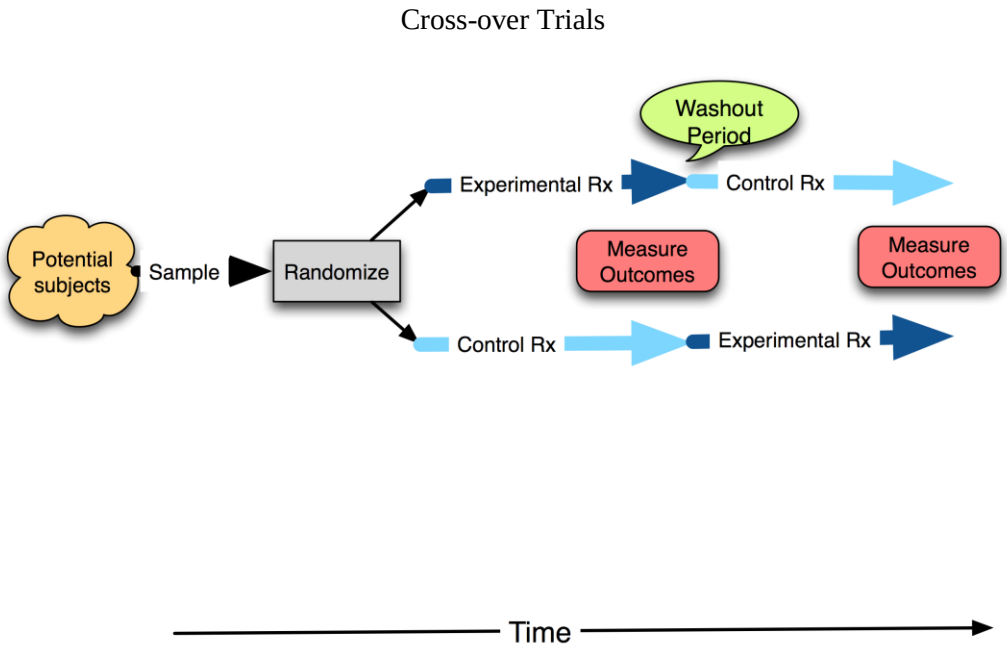
When possible, RCTs in rare diseases should be designed with an active comparator group in order to enhance recruitment.[20] As an example of an active comparator, the single RCT of factor replacement prophylaxis in hemophilia A, a rare sex-linked genetic disorder, compared prophylactic factor VIII infusions against an enhanced episodic factor VIII infusion schedule.[21] Although the enhanced episodic infusion of factor VIII had never been investigated in a placebo-controlled trial, it was considered to be the most reasonable standard of therapy; the investigators considered it to have been unethical to randomize subjects to a placebo.

Post-trial provision of beneficial treatment / open-labeled extension

Patients may be more willing to participate in trials if they are offered access to the experimental treatment after the trial end (no matter which treatment arm they had been assigned to). There is an ethical support for this; the Declaration of Helsinki states that at the end of the study patients should be informed about the results, and share any benefits from the trial findings (such as having access to beneficial interventions established by the study).[22] In fact, many national and international guidelines support post-trial mandatory provision of beneficial interventions.[23] The interruption of treatment at the end of the trial can create frustration and a feeling of exploitation among participants. In contrary, when the experimental treatment is offered at the end of trial for as long as necessary, patients may feel rewarded for their participation. Moreover, they may agree to a continued follow-up that will provide data for a longer-term effectiveness and safety (often called an ‘open-labeled extension phase’). Some trial sponsors argue that a post-trial provision may be very costly and may take funds from other potential projects;[24] however, in rare diseases the overall budget impact is likely to be small because the number of eligible patients is small as well. Post-trial provisions may encourage patients to participate in RCTs and help investigators to fulfill their ethical obligations.

FIG. 2 Cross-over Study

In the cross-over design, subjects are randomized to a first period in which some start with experimental treatment, and the others with control treatment. At the end of the first period outcomes are measured. If required, there is then a washout period after which each subject is exposed to the other treatment group. At the end of the second period outcomes are measured again.



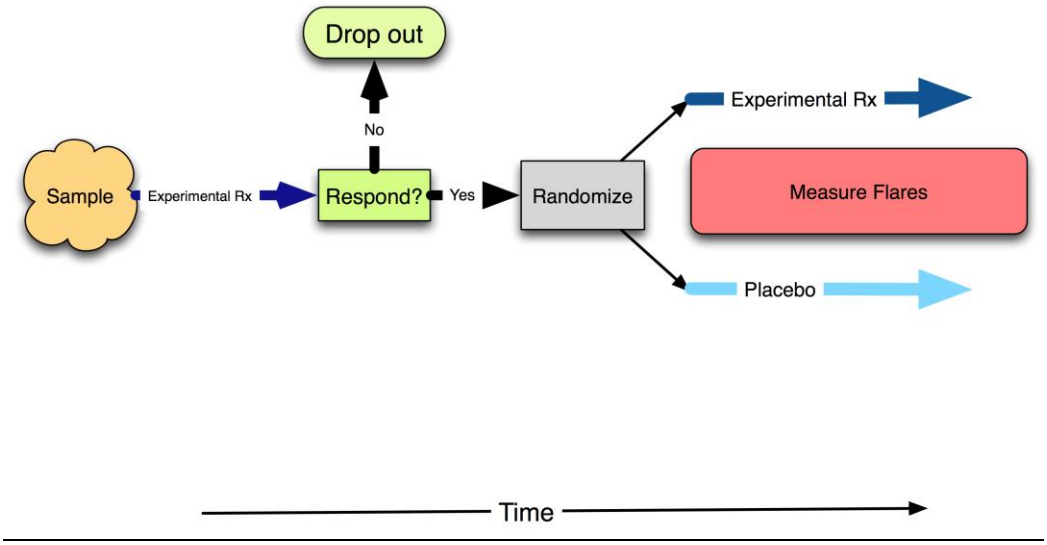
Cross-over trials compare two (or more) treatments by allocating each participant to all compared treatments in a randomly selected sequence. The design offers many advantageous features to investigators of rare diseases such as less variability (as each patient acts as his or her own control allowing for within patient comparisons) / higher precision – and therefore the need for a smaller sample size (in some cases almost half of what is needed for a parallel design). In addition all patients will receive the experimental treatment at some time during the trial which may lead to enhanced acceptability and improved recruitment.[25,26] The design is limited to studies of chronic, stable conditions, to investigation of symptom relief, to short-term therapies, and for treatments that do not induce a permanent effect.

Cross-over trials have been widely used in pediatrics, cancer research, clinical pharmacology and psychiatry.[26] For example, a review of all RCTs in the Archives of Disease in Childhood from 1982–1996 found that about one-third used a cross-over design.[27] Methodological challenges that need careful consideration in cross-over trials include an effective washout period (time necessary to wait before administrating the next medication to clear the effect of the prior medication, to avoid a carryover effect), order effect and period effect.[28] The design is also more sensitive to dropouts and missing data as each patient carries more weight in the analysis.

Enriched enrolment, randomized withdrawal The enriched enrolment randomized withdrawal (EERW) design addresses ethical concerns and patient preferences about placebo assignment by shortening the time patients are on control therapy/placebo. The EERW trial design has two phases and was first described in 1975.[29] The first phase, *the enrichment phase,* is used to identify the experimental drug responders by enrolling study participants into an open-label trial where all patients receive the treatment under the study. Once the responders are identified, they are enrolled into the *randomized withdrawal phase* where they are randomly allocated to continue receiving experimental treatment or switch to a control treatment. The trial endpoints are usually the return of symptoms.[30] The enrichment design effectively increases the average benefit of the experimental treatment over the control. Since even those subjects who have therapy withdrawn can re-start effective experimental therapy when they have flared, the time on placebo is limited. This is thought to increase acceptability and accrual; as such, this design has been widely used, e.g., in pediatric rheumatology trials.[31]

FIG. 3 Randomized withdrawal design

In this design, all subjects are begun on the experimental therapy. Those that respond are then randomized to continue experimental therapy, or to be withdrawn to placebo. The groups are followed to see which group has a greater flare rate.



A non-exhaustive, keyword search 'randomized withdrawal' conducted in the MEDLINE and EMBASE databases from January 2000 until April 2011 (limiting to English language and human studies) identified 42 abstracts, 17 of which described original RCTs with EERW design. The review of these abstracts revealed that almost half were related to pain management and one-third to psychiatric conditions. All 17 studies used placebo as a control and only 5 used a time-to-event outcome as a primary end-point. This review also indicated that the use of EERW design is increasing as 11 of 17 studies were conducted after 2008.

The methodological challenges of this design include potential carryover effects, establishment of the enrichment duration, disease activity status ascertainment (assessment if this is a real remission/improvement), data imputation methods for missing values for withdrawals (unless time-to-event analyses are applied) and capturing of long-term adverse events.[30]

Randomized placebo-phase design

The Randomized placebo-phase design (RPPD) was developed with the primary aim to improve the acceptability of entering a trial by effectively decreasing the time necessary to be allocated to placebo.[32] It was designed for treatments that may produce a lasting remission or response (and so would not be suitable for cross-over designs). As in the parallel group RCT (Figure 1), patients in the RPPD are first randomly allocated to either an experimental or control group. However, after a short, fixed time period (called the placebo-phase) patients in the control group are blindly switched to the experimental treatment (Figure 4). The design is based on the assumption that if the treatment is effective, patients who receive it sooner will respond, on average, sooner. At the end of the trial, average response times of the groups are compared, most often using time-to-event analysis.

FIG. 4 Randomized placebo phase design (RPPD)

In the RPPD, subjects are randomized to begin therapy earlier, or later. Those that start later are begun on a period of placebo to preserve blinding. Outcomes are measured as time-to-response

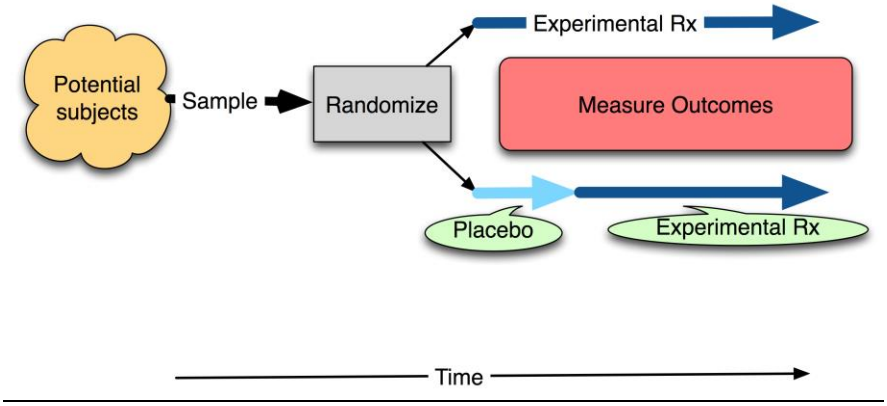
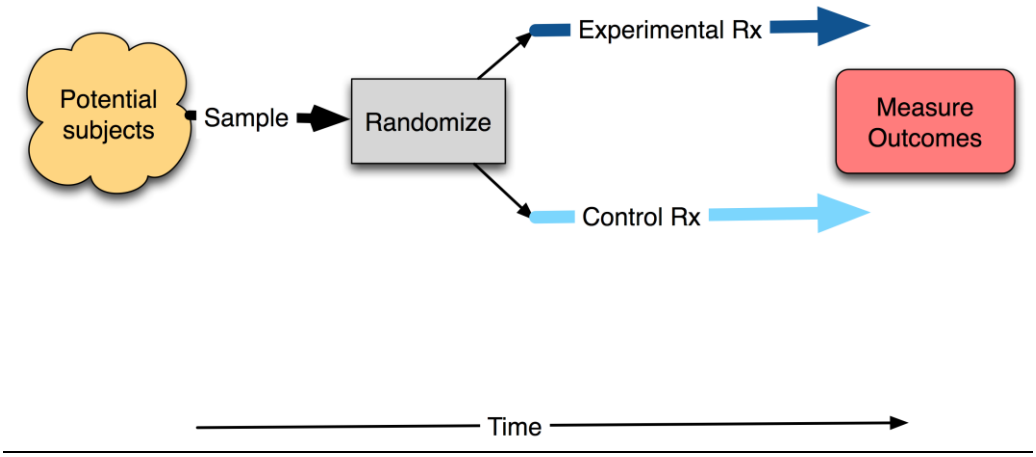


FIG. 1 Parallel groups randomized controlled trial (RCT)

In the RCT, subjects are randomized to take either experimental treatment or control treatment (often placebo). Outcomes may be measured at the end of the study period, or during the study period (longitudinally, or as time-to-event).



If the RPPD is, in fact, more acceptable to potential subjects and enrolling physicians, there is a real benefit to the design. We did a computer simulation study assessing the performance of the RPPD against the parallel group RCT under different time-to-response distributions.[33] The results revealed that under the lognormal and Weibull distributions the RPPD can be very close in efficiency (statistical power) to the RCT. If the RPPD makes it easier to accrue subjects, there is a real benefit for rare diseases. Another computer simulation study established effective internal monitoring and early stopping rules for the RPPD.[34]

Response time distributions for specific patient population and treatment modality can be estimated from patient registries or past clinical trials.[35] The establishment of the optimal placebo-phase duration and overall trial duration is very important. A longer placebo-phase duration will decrease the required sample size; however, the optimal duration is defined only after carefully considering patient preferences, the disease progress model, the minimal clinically important drug potency / effect size, and the response time to drug statistical distribution.[33]

Use Bayesian adaptive trial design and analysis to achieve interpretable results from small studies

Bayesian analysis allows us to provide estimates of treatment effectiveness even in small studies.

The traditional, or frequentist, paradigm for the analysis of clinical trials focuses on the probability (p-value) that the observed data (or values more extreme) were obtained by chance alone (i.e. under the null hypothesis).[36] By convention, when the probability is less than 0.05, we conclude that the observed difference is "statistically significant". The 0.05 probability of incorrectly concluding that there is a treatment effect (if in fact there is none) is known as a Type I (or false positive) error.[37] Therefore, we interpret the question answered by the p-value as only whether there is a treatment effect, rather than how large that treatment effect may be. Estimates of the magnitude of treatment effects are however of most relevance to decision makers.[38] The magnitude of the p-value is not only related to the effect size, but also the sample size (which helps determine the precision of the study).[36] As such, the traditional approach is not ideally suited to the study of rare diseases, where larger clinical trials may not be feasible.

In contrast, a Bayesian analysis allows for maximal information to be gained from a limited number of subjects. The Bayesian approach achieves this, in part, by providing for formal incorporation of prior information into the analysis, which may reduce sample size requirements.[39]A Bayesian analysis also provides a meaningful estimate of the direct probability of any given magnitude of treatment response.

Bayesian approaches to assessment of a health technology can be defined as an "explicit and qualitative use of external evidence in the design, monitoring, analysis, interpretation and reporting of the results".[40] In addition to analysis, the Bayesian approach can be used in the design and monitoring of clinical trials, including sample size estimation and stopping rules.[40]

Central to this approach, is the notion of the *prior probability distribution*, which is a probability distribution of the variable of interest (say, the effect size of an experimental treatment) from data external to the study. This prior may be based on previous studies or on expert opinion.[41] A Bayesian analysis is a formal method of integrating this prior distribution with the distribution of the new data – known as the *likelihood*, which is the probability of the data observed, or more extreme, given a hypothesis (see above for a description of p values) – to yield a *posterior probability distribution*.[40] For example, in the analysis of a clinical trial of an experimental treatment, the prior distribution may be an estimate of the treatment response from previous studies (early phase studies, or studies in other populations). This is then combined with the actual data from the trial to yield a posterior distribution, which is the new, and better, estimate of the probability of the treatment response.

Formal incorporation of data from previous studies, may allow investigator to address questions meaningfully with fewer subjects;[39] however, inclusion of prior information is controversial. Some suggest that we can't be certain that the information contained in the prior

is correct.[42] In order to address this criticism, Bayesian analyses can be done with a specific type of prior – a skeptical prior – which is a prior that is constructed to have a high probability of there *not* being an experimental treatment effect.[40] An analysis using a skeptical prior are particularly important for clinical trials. Since the skeptical prior reflects a low probability of a meaningful treatment response—and the posterior distribution updates this probability on the basis of the data—this posterior probability distribution, therefore, is what an individual who holds a skeptical view of the treatment should believe with new knowledge from the trial data. Using a skeptical prior, we ask whether the data from our trial are sufficiently strong to convince a skeptic to adopt the treatment.

A major benefit for the study of rare disease is that, unlike the traditional method, the Bayesian method yields a direct estimate of the probability of a meaningful treatment effect.[43] This is true, no matter what the size of the data collection is. As is true for the traditional frequentist approach, more data will yield a more precise estimate, but with the Bayesian approach we are not limited by arbitrary decisions regarding "significance" based on a p value.

This can be illustrated with data from a hypothetical placebo controlled trial with 30 participants (15 in each group) of a drug for symptom management of a rheumatologic disorder. The outcome of this study is a quality of life instrument. A difference of 2-points is believed to be clinically important. The mean value in the experimental arm is 13.6 units (standard deviation: 4.97) and in the control arm 10.5 units (standard deviation: 4.36). A traditional frequentist analysis with a t-test provides insufficient evidence to disprove the null hypothesis – t-statistic 1.797, df = 28, p = 0.083, mean difference 3.06: 95% confidence interval: -0.42 to 6.56. For this small study, the power is low and the results may be false negative. A Bayesian analysis (using an uninformative prior[1]) provides very similar estimates for the difference in means between the groups (median value: 3.05, 95% credible interval -0.58 to 6.61). However, unlike the frequentist analysis the Bayesian analysis allows us to determine that the probability of a clinically important difference of 2 points is 74% – which can be thought of as 3:1 odds favoring the experimental treatment. For an inexpensive and safe treatment, this may be enough evidence to support treatment. There is a 96% probability that the experimental treatment is at least a little bit better. Therefore, this small study, that would have likely been regarded as "negative" with the frequentist approach, may provide useful information when analyzed by the Bayesian approach. A Bayesian reanalysis of a small RCT of methotrexate in scleroderma came to similar conclusions.[44]

Get more information from individual subjects

Multiple cross-over designs

As discussed above, in a cross-over study each subject's experience on both experimental and control therapy is contrasted. Because the comparison is 'within subject', the variability in response is reduced (i.e. the precision is increased) and fewer subjects are needed to demonstrate an important effect size.[45] In a multiple cross-over study, each subject is crossed between experimental and control treatments several times; in this way, we gain additional information for each subject, and the total number of subjects in a study may be decreased without sacrificing power.[46]

One problem with multiple cross-over designs is that they are administratively complex, and they are very sensitive to drop-out. If a subject drop-outs from one of the treatment periods, all that subject's data must be dropped from a traditional analysis, or the analysis has to be changed to account for the drop-out.

Multiple n-of-1 trials

A more flexible method is to combine n-of-1 trials using a form of Bayesian meta-analysis.[47] N-of-1 studies are randomized clinical trials, with multiple cross-overs, in single subjects.[48,49] Because cross-over periods are assigned, the design is limited by the same factors that limit other crossover designs (see above). N-of-1 trials

[1]An uninformative prior probability distribution assumes that we have no knowledge of the treatment effect before we start our study. It is therefore uncontroversial, but doesn't allow us to take full advantage of the Bayesian approach.

are flexible by nature, since the number of periods may be varied according to each subject; the study can be continued until a definitive conclusion can be made for that subject being studied. However, because they are studies done in a single subject, no generalizations can be made about the studied treatment to others.

By combining subjects who have had n-of-1 studies (in each of whom there is a precise understanding of how well the experimental treatment works) it is possible to estimate the treatment effect in the population. For example, we studied the anti-emetic effects of metopimazine combination therapy in children getting chemotherapy for brain tumours,[50] the pain relieving effects of amitriptyline in children with arthritis,[51] and the effect of topical vitamin E in preventing mucositis in children receiving cancer chemotherapy,[52] using this method. Each study was small (total sample size ranged from 6 to 16 subjects) yet each provided us with an estimate of treatment effectiveness adequate to draw conclusions about its use. We used a Bayesian meta-analytic technique to combine the subjects' data, but other statistical methods may also be used.[53] In situations in which a crossover design may be used, and maximal flexibility is desired, multiple n-of-1 studies may be a helpful strategy.

Use adjustment to get unbiased estimates from observational data

The rigorous use of already collected treatment data (observational data) – say, from clinical charts – is an alternative solution. There is much useful data about treatment effectiveness for rare diseases that lives in our existing records. As discussed above, though, confounding by indication is a critical threat to the validity of inferences made using this approach, and as such, must be accounted for. Confounding by indication occurs when there is non-comparability between the study groups resulting from the way they were constructed. Exposed and unexposed patients may differ systematically in important characteristics. That is, there is a reason why some patients are treated with one therapy, and other patients get another – they may be more severe, have a poorer prognosis, have different health care access, etc. It is often these patient differences that have more to do with outcome than the treatments themselves. Small differences in many covariates can accumulate into substantial overall differences.[54] This can result in biased estimation of treatment effect.

However, innovative study design and analytic techniques can be used to make the study groups comparable. In this way, unbiased estimates of treatment effect can be achieved that are comparable to similarly sized RCTs. One strategy is the use of propensity score (PS) methods. A propensity score is the conditional probability of assignment to a treatment based on the observed covariates.[55] For each individual, the PS is a measure of the likelihood that a person would have been treated with a particular treatment using their baseline characteristics.[56] It reduces the collection of baseline characteristics to a single composite score that appropriately summarizes the collection of characteristics.[54] Once estimated, the PS can be used to create comparable groups through pair matching on the PS, sub-classification on the PS or covariate adjustment using the PS.[55,57] Therefore, the PS can be thought of as a balancing score, to create groups that are comparable. The test of a good PS model is the degree to which it results in the baseline characteristics being balanced between exposed and unexposed individuals.[58]At any value of the PS, the difference between the treatment and control groups is an unbiased estimate of the average treatment effect.

The use of PS methods has advantages compared to other adjustment methods. PS are more reliable tools because the assumptions needed to make their answers appropriate are transparent.[54] The methods also allow for inspection of the data to assess whether the exposed and unexposed groups overlap enough in characteristics to allow sensible estimation of treatment effect.[54] In the setting of rare disease, PS methods confer additional value. Commonly used regression techniques typically require at least 10 outcomes for every covariate included in a model. In the setting of a rare disease, there may not be sufficient data for necessary covariates to be included. By reducing the important baseline covariates to a single composite score (the PS),

the necessary covariates can be accounted for using data from a small sample of patients.

We have used PS methods to evaluate the effect of warfarin on survival in scleroderma associated pulmonary arterial hypertension (SSc-PAH).[59] SSc-PAH is an uncommon disease with an estimated prevalence of 2.93 per million.[60] SSc-PAH has a poor prognosis, with a median survival rate as low as 12 months.[61]Anticoagulation with warfarin is recommended to improve survival in these patients. However there is no evidence to support this recommendation. We evaluated the ability of warfarin to improve survival in these patients using observational data. We found the SSc-PAH patients exposed to warfarin had more severe disease and used more PAH-specific medications than warfarin unexposed patients. Thus the crude association between survival and warfarin use is likely to be confounded (confounding by indication). If these differences were not accounted for, the estimated treatment effect of warfarin on survival would be biased. This would have led to the conclusion that warfarin worsens survival. Bayesian propensity scores were used to adjust for differences between patients exposed and not exposed to warfarin, and assemble a matched cohort. In the matched cohort, the hazard ratio was 1.06 and the probability that warfarin improves median survival by 6-months or more is only 23.5%. We concluded that there is a low probability that warfarin improves survival by an important amount in SSc-PAH. Our use of propensity score matching reduced the effect of confounding by indication allowing us to make a less biased estimate of treatment effect.

PS matching has been applied to the study of early aggressive corticosteroid treatment in juvenile dermatomyositis – a commonly practiced treatment strategy for which a RCT has never been done because of the rarity of the disorder.[62] More complicated models – including marginal structural modeling[63] and instrumental variable modeling[64] – may be useful in situations that are more complex. For example, we used a marginal structural model when evaluating the effectiveness of intravenous immunoglobulin for resistant juvenile dermatomyositis;[65] this is another widely used treatment that could not have been studied by an RCT given the rarity of the disorder.

SUMMARY

Rare disorders pose challenges to clinicians and scientists who are interested in applying the best evidence-based treatment decisions to medical care. While the parallel groups RCT is considered the gold standard experimental method, there are situations in which RCTs cannot feasibly be done. The solution proposed is for the investigator to consider a flexible "toolkit" of study designs that can be applied to generate the needed treatment evidence. This toolkit – which includes designs to maximize the acceptability of studies, get more information from fewer subjects, use more flexible and informative analytic methods, and derive valid treatment conclusions from already collected data – is only limited by the imagination of investigators.

Acknowledgements

Lusine Abrahamyan is supported by a Canadian Cardiovascular Outcomes Research Team postdoctoral award.

Ivan Diamond was supported by a Graduate Studentship Award from the Canadian Liver Foundation and the Chisholm Memorial Fellowship, Post Graduate Medical Education Office, the University of Toronto, with additional support from the Surgeon Scientist Training Program, Department of Surgery, University of Toronto.

Sindhu Johnson is supported by a Canadian Institutes of Health Research Clinician Scientist Award and the Norton-Evans Fund for Scleroderma Research.

Brian Feldman is supported by The Ho Family Chair in Autoimmune Diseases.

REFERENCES

1. Institute of Medicine. Rare disease and orphan products: Accelerating research and development. Washington, D.C.: The National Academies Press. 2010;420.
2. Gustason W. Reasoning from evidence. Inductive logic. New York: Macmillan College Publishing Company. 1994;318.

3. Morabia APCA. Louis and the birth of clinical epidemiology. J Clin Epidemiol 1996;49(12):1327-33.
4. Wikipedia contributors. Evidence-based medicine: Wikipedia, The Free Encyclopedia.; 2011 [updated 27 April 2011 08:08 UTC; cited 2011 12 May]. Available from: http://en.wikipedia.org/w/index.php?title=Evidence-based_medicine&oldid=426171031.
5. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. BMJ 1996;312(7023):71-2.
6. Savitz DA. Interpreting Epidemiologic Evidence. Strategies for study design and analysis. Oxford: Oxford University Press, Inc.; 2003.
7. Feinstein AR. Clinical Epidemiology. The Architecture of clinical research. Philadelphia: W. B. Saunders Company; 1985.
8. Bosco JL, Silliman RA, Thwin SS, et al. A most stubborn bias: no adjustment method fully resolves confounding by indication in observational studies. J Clin Epidemiol. 2010;63(1):64-74.
9. Kaplan W, Laing R. Priority medicines for Europe and the World. World Health Organization, Department of Essential Drugs and Medicines Policy. 2004. Available: http://mednet3.who.int/prioritymeds/repo?rt/final18october.pdf. Accessed 23 April, 2011.
10. Hughes DA, Tunnage B, Yeo ST. Drugs for exceptionally rare diseases: do they deserve special status for funding? QJM 2005;98(11):829-36.
11. European Organisation for Rare Diseases (EURORDIS). Rare diseases: understanding this public health priority. 2005. Available: http://www.eurordis.org/IMG/pdf/princeps_document-EN.pdf. Accessed 23 April, 2011.
12. Gerss JWO, Kopcke W. Clinical trials and rare diseases. Adv Exp Med Biol 2010;686:173-90.
13. Behera M, Kumar A, Soares HP, Sokol L, Djulbegovic B. Evidence-based medicine for rare diseases: implications for data interpretation and clinical trial design. Cancer Control 2007;14(2):160-6.
14. Gallin JI, Alling DW, Malech HL, et al. Itraconazole to prevent fungal infections in chronic granulomatous disease. N Engl J Med 2003;348(24):2416-22.
15. Wilcken B. Rare diseases and the assessment of intervention: what sorts of clinical trials can we use? J Inherit Metab Dis 2001;24(2):291-8.
16. Abrahamyan L, Johnson SR, Beyene J, Shah PS, Feldman BM. Quality of randomized clinical trials in juvenile idiopathic arthritis. Rheumatology (Oxford) 2008;47(5):640-5.
17. Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. N Engl J Med 1988;318(26):1728-33.
18. Halpern SD, Karlawish JH, Casarett D, Berlin JA, Townsend RR, Asch DA. Hypertensive patients' willingness to participate in placebo-controlled trials: implications for recruitment efficiency. Am Heart J 2003;146(6):985-92.
19. Halpern SD, Ubel PA, Berlin JA, Townsend RR, Asch DA. Physicians' preferences for active-controlled versus placebo-controlled trials of new antihypertensive drugs. J Gen Intern Med 2002;17(9):689-95.
20. Makuch RW, Johnson MF. Dilemmas in the use of active control groups in clinical research. IRB 1989;11(1):1-5.
21. Manco-Johnson MJ, Abshire TC, Shapiro AD, et al. Prophylaxis versus episodic treatment to prevent joint disease in boys with severe hemophilia. N Engl J Med 2007;357(6):535-44.
22. The World Medical Association Declaration of Helsinki. Ethical Principles for Medical Research Involving Human Subjects. 59th WMA General Assembly. Seoul: 2008 October. Report No.
23. Zong Z. Should post-trial provision of beneficial experimental interventions be mandatory in developing countries? J Med Ethics 2008;34(3):188-92.
24. Grady C. The challenge of assuring continued post-trial access to beneficial treatment. Yale J Health Policy Law Ethics 2005;5(1):425-35.
25. Senn S. The AB/BA crossover: past, present and future? Stat Methods Med Res 1994;3(4):303-24.
26. Elbourne DR, Altman DG, Higgins JP, Curtin F, Worthington HV, Vail A. Meta-analyses involving cross-over trials: methodological issues. Int J Epidemiol 2002;31(1):140-9.
27. Campbell H, Surry SA, Royle EM. A review of randomised controlled trials published in Archives of Disease in Childhood from 1982-96. Arch Dis Child 1998;79(2):192-7.
28. Reed JF, 3rd. Analysis of two-treatment, two-period crossover trials in emergency medicine. Ann Emerg Med 2004;43(1):54-8.
29. Straube S, Derry S, McQuay HJ, Moore RA. Enriched enrollment: definition and effects of enrichment and dose in trials of pregabalin and

gabapentin in neuropathic pain. A systematic review. Br J Clin Pharmacol 2008;66(2):266-75.
30. Katz N. Enriched enrollment randomized withdrawal trial designs of analgesics: focus on methodology. Clin J Pain 2009;25(9):797-807.
31. Lovell DJ, Giannini EH, Reiff A, et al. Etanercept in children with polyarticular juvenile rheumatoid arthritis. Pediatric Rheumatology Collaborative Study Group. The New England journal of medicine 2000;342(11):763-9.
32. Feldman B, Wang E, Willan A, Szalai JP. The randomized placebo-phase design for clinical trials. J Clin Epidemiol 2001;54(6):550-7.
33. Abrahamyan L, Li CS, Beyene J, Willan AR, Feldman BM. Survival distributions impact the power of randomized placebo-phase design and parallel groups randomized clinical trials. J Clin Epidemiol 2010;64(3):286-92.
34. Shook S. Randomized placebo-phase design: evaluation, interim monitoring and analysis. Diss: University of Pittsburgh; 2010.
35. Abrahamyan L, Beyene J, Feng J, et al. Response times follow lognormal or gamma distribution in arthritis patients. J Clin Epidemiol 2010;63(12):1363-9.
36. Goodman SN. Toward evidence-based medical statistics. 1: The P value fallacy. Ann Intern Med 1999;130(12):995-1004.
37. Sterne JA, Davey Smith G. Sifting the evidence-what's wrong with significance tests? BMJ 2001;322(7280):226-31.
38. Lilford RJ, Thornton JG, Braunholtz D. Clinical trials and rare diseases: a way out of a conundrum. BMJ 1995;311(7020):1621-5.
39. Diamond GA, Kaul S. Prior convictions: Bayesian approaches to the analysis and interpretation of clinical mega trials. J Am Coll Cardiol 2004;43(11):1929-39.
40. Spiegelhalter DJ, Myles JP, Jones DR, Abrams KR. Bayesian methods in health technology assessment: a review. Health Technol Assess 2000;4(38):1-130.
41. Johnson SR, Tomlinson GA, Hawker GA, Granton JT, Feldman BM. Methods to elicit beliefs for Bayesian priors: a systematic review. J Clin Epidemiol 2010;63(4):355-69.
42. Howard G, Coffey CS, Cutter GR. Is Bayesian analysis ready for use in phase III randomized clinical trials? Beware the sound of the sirens. Stroke 2005;36(7):1622-3.
43. Goodman SN. Toward evidence-based medical statistics. 2: The Bayes factor. Ann Intern Med 1999;130(12):1005-13.
44. Johnson SR, Feldman BM, Pope JE, Tomlinson GA. Shifting our thinking about uncommon disease trials: the case of methotrexate in scleroderma. J Rheumatol 2009;36(2):323-9.
45. Menzies D, Popa J, Hanley JA, Rand T, Milton DK. Effect of ultraviolet germicidal lights installed in office ventilation systems on workers' health and wellbeing: double-blind multiple crossover trial. Lancet 2003;362(9398):1785-91.
46. Gilron I, Booher SL, Rowan JS, Max MB. Topiramate in trigeminal neuralgia: a randomized, placebo-controlled multiple crossover pilot study. Clin Neuropharmacol 2001;24(2):109-12.
47. Zucker D, Schmid C, McIntosh M, D'Agostino R, Selker H, Lau J. Combining single patient (N-of-1) trials to estimate population treatment effects and to evaluate individual patient responses to treatment. J Clin Epidemiol 1997;50(4):401-10.
48. Guyatt G, Sackett D, Adachi J, et al. A clinician's guide for conducting randomized trials in individual patients. CMAJ 1988;139(6):497-503.
49. Guyatt GH, Heyting A, Jaeschke R, Keller J, Adachi JD, Roberts RS. N of 1 randomized trials for investigating new drugs. Control Clin Trials 1990;11(2):88-100.
50. Nathan PC, Tomlinson G, Dupuis LL, et al. A pilot study of ondansetron plus metopimazine vs. ondansetron monotherapy in children receiving highly emetogenic chemotherapy: a Bayesian randomized serial N-of-1 trials design. Support Care Cancer 2006;14(3):268-76.
51. Huber AM, Tomlinson GA, Koren G, Feldman BM. Amitriptyline to relieve pain in juvenile idiopathic arthritis: a pilot study using Bayesian meta-analysis of multiple N-of-1 clinical trials. J Rheumatol 2007;34(5):1125-32.
52. Sung L, Tomlinson GA, Greenberg ML, et al. Serial controlled N-of-1 trials of topical vitamin E as prophylaxis for chemotherapy-induced oral mucositis in paediatric patients. Eur J Cancer 2007;43(8):1269-75.
53. Zucker DR, Ruthazer R, Schmid CH. Individual (N-of-1) trials can be combined to give population comparative treatment effect estimates: methodologic considerations. J Clin Epidemiol 2010;63(12):1312-23.
54. Rubin DB. Estimating causal effects from large data sets using propensity scores. Ann Intern Med 1997;127(8 Pt 2):757-63.

55. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika 1983;70(1):41-55.
56. D'Agostino RB. Tutorial in biostatistics. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. Statist Med 1998;17:2265-81.
57. Rosenbaum PR, Rubin DB. Reducing bias in observational studies using subclassification on the propensity score. JAMA 1984;79(387):516-24.
58. Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. Statist Med 2009;28:3083-107.
59. Johnson SR, Granton JT, Tomlinson GA, et al. Warfarin in scleroderma-associated and idiopathic pulmonary arterial hypertension. A Bayesian approach to evaluating treatment in uncommon disease. Submitted 2011.
60. Condliffe R, Kiely DG, Peacock AJ, et al. Connective tissue disease-associated pulmonary arterial hypertension in the modern treatment era. Am J Respir Crit Care Med 2009;179(2):151-7.
61. Koh ET, Lee P, Gladman DD, Abu-Shakra M. Pulmonary hypertension in systemic sclerosis: an analysis of 17 patients. Br J Rheumatol 1996;35(10):989-93.
62. Seshadri R, Feldman BM, Ilowite N, Cawkwell G, Pachman LM. The role of aggressive corticosteroid therapy in patients with juvenile dermatomyositis: a propensity score analysis. Arthritis Rheum 2008;59(7):989-95.
63. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology 2000;11(5):550-60.
64. Stukel TA, Fisher ES, Wennberg DE, Alter DA, Gottlieb DJ, Vermeulen MJ. Analysis of observational studies in the presence of treatment selection bias: effects of invasive cardiac management on AMI survival using propensity score and instrumental variable methods. JAMA 2007;297(3):278-85.
65. Pullenayegum EM, Lam C, Manlhiot C, Feldman BM. Fitting marginal structural models: estimating covariate-treatment associations in the reweighted data set can guide model fitting. J Clin Epidemiol 2008;61(9):875-81.