A01

INTERIM EVALUATION OF EFFICACY OR FUTILITY IN GROUP-SEQUENTIAL TRIALS WITH TWO TIME-TO-EVENT OUTCOMES

TOSHIMITSU HAMASAKI

NATIONAL CEREBRAL AND CARDIOVASCULAR CENTER

We discuss logrank test-based methods for efficacy or futility evaluation in group-sequential clinical trials that compare two interventions with respect to two time-to-event outcomes. Evaluation is conducted under three situations: (a) both events are non-composite and non-fatal, (b) both events are non-composite, but one event is fatal, and (c) one event is composite, but other is fatal and non-composite. Based on group-sequential boundaries, we consider several decision-making frameworks for evaluating efficacy or futility. We consider two inferential goals, evaluating if a test intervention is superior to a control intervention on: (i) both outcomes (Co-primary endpoints: CPE), and (ii) at least one outcome (multiple primary endpoints: MPE). For the CPE goal, we incorporate the correlations among the outcomes into the calculations for non-binding futility boundaries and sample sizes (or event numbers) as a function of other design parameters, including mean differences, the number of analyses, and efficacy boundaries. We investigate the operating characteristics of the decision-making frameworks in terms of efficacy/futility boundaries, power, the Type I error rate, sample sizes, event numbers, while varying the number of analyses, the correlations among the outcomes, and hazard ratios. We provide examples to illustrate the methods and discuss practical considerations when designing group-sequential designs in clinical trials with two time-to-event outcomes.

Contributors

Koko Asakura
National Cerebral and Cardiovascular Center

Scott Evans
Harvard T.H.Chan School of Public Health

Tomoyuki Sugimoto
Kagoshima University

Chin-Fu Hsiao
National Health Research Institutes

Haruko Yamamoto
National Cerebral and Cardiovascular Center

A02

COLLECTION OF REAL-TIME DATA FROM PARTICIPANTS – INTERACTIVE TEXT SURVEYS

GEORGE KITSARAS

UNIVERSITY OF MANCHESTER

Introduction. Questionnaire-based assessments are associated with challenges such as recall bias and low response and retention rates. Real-time, mobile-based approaches in behavioural, psychological and social research can present a new way forward. An area where those real-time mobile-based approaches can be valuable are bedtime routines, an area associated with child wellbeing and development. Bedtime routines have been difficult to evaluate with traditional assessments such as questionnaires mainly due to their retrospective approach. Approximately 95% of the population in the United Kingdom and the United States owns a mobile phone, sending an average of 100 text messages per subscription per month. Moreover, some hard to reach populations, such as low-income and minority groups, not only show similar rates of using mobile phones but they also report higher percentage of text-messaging than other groups. Methods. Preliminary Patient and public involvement (PPI) work with families showed a clear preference for a text-survey assessment of bedtime routines via an interactive text survey. In collaboration with a company specializing in text-surveys, an interactive, user-friendly, real time text-survey assessment of bedtime routines was developed and administered to 50 families with preschool age children (ages 3-6). The assessment was delivered to participating parents' mobile phones for 5 consecutive nights, after the child had been to bed and it involved open-ended and closed questions about that night's bedtime routine. There was a maximum of 10 and a minimum of 8 questions depending on response patterns. The average completion time was 2-minutes. All data points were managed and analysed electronically via a secure online platform. Anonymised feedback, response and retention rates and other insight information were collected. Results. The text survey was perceived positively with an average score of 4.5 out of 5 for overall experience. There was an overall response rate of 87%, much higher than conventional questionnaire-based assessments. Finally, retention rates were good with every participant replying to the text-survey at least 3 out of 5 nights resulting in an average of 40 unique data points per participant throughout the duration of the study. There were no dropouts during the study. Conclusion. Text-survey assessment delivered to participant's mobile phones was successful for assessing different types of bedtime routines. The assessment was perceived positively causing minimum disruption while obtaining extensive real-time data. Traditional questionnaire-based approaches are limited in the depth and quality of data they can gather while interactive text-surveys offer great potential for real time assessment of general and health-related behaviours. Finally, mobile text messaging-based approaches can have intervention potential in terms of tailored, personalised, on-demand and real-time support for bedtime routines and health behaviours alike.

Contributors

Michaela Goodwin
Julia Allan
Michael KellyIain Pretty

A03

MONITORING INTERACTIVE RESPONSE TECHNOLOGY VENDOR IMPLEMENTED RANDOMIZATION AND DOSING SYSTEMS

JASON MCCOY

AMGEN, INC.

Use of Interactive Response Technology (IRT) vendor-implemented randomization and dosing algorithms is an effective method to reduce the incidence of medication errors in the clinical trial setting. However, errors in the programming of randomization and dosing algorithms remains a risk factor leading to medication errors for subjects on a clinical trial. Hence, vendor oversight is important in reducing risks for subjects participating in clinical trials. Additionally, regulatory agencies have maintained that it is ultimately the sponsor's responsibility to maintain oversight of IRT vendor's systems used in their clinical trials. We describe a method for improved monitoring of Interactive Response Technology vendor implemented randomization and dosing systems shortening detection time and reducing impact of system errors. To augment best practice IRT system testing and vendor oversight activities, we have developed a process which enables a sponsor to continuously monitor randomization and dosing algorithms such that the sponsor will become aware of potential IRT-caused dosing discrepancies within 24 hours of first occurrence, limiting the potential for a systematic error to impact a large number of subjects. Implementing this process represents an important step in ensuring subject safety and the validity of clinical trials.

Contributors

Venkat Mandarapu

A04

SAMPLE SIZE ESTIMATION USING A HYBRID CLASSICAL AND BAYESIAN PROCEDURE

MARIA CIARLEGLIO

YALE SCHOOL OF PUBLIC HEALTH

The hypothesized treatment effect and estimated nuisance parameters play an important role in a study's sample size and power calculation. Point estimates for these parameters are often calculated using historical data. However, the uncertainty in these estimates is rarely addressed. We present a hybrid classical and Bayesian procedure that formally integrates prior information on the distributions of the hypothesized study parameters into the study's power calculation. Conditional expected power, which averages the traditional power curve using the prior distributions of the unknown parameters as the averaging weight, is used, and the sample size is found that equates the pre-specified frequentist power and the conditional expected power of the trial. A method for sample size re-estimation in which the prior distributions are updated using partial trial data is also discussed. We show that using the proposed method for sample size determination during the design phase helps to protect against misspecification of the nuisance parameters, reducing the probability of an underpowered study.

Contributors

A05

TWO FACTORS IMPACTING FORMULATION OF DATA MONITORING COMMITTEE RECOMMENDATIONS FOR ADAPTIVE DESIGN CLINICAL TRIALS

NAVNEET HAKHU

AXIO RESEARCH, LLC

Data Monitoring Committees (DMCs) are charged with monitoring ongoing clinical trial(s) to protect the interests of participating human volunteers and preserve trial integrity through interim reviews of safety and efficacy data. The DMC's assessment of risk-benefit forms the basis of the committee's recommendations to the sponsor. Pre-specified monitoring guidelines (also called stopping boundaries) may be incorporated into an adaptive component of a trial's design for efficacy, futility, and/or safety/harm, which can facilitate the DMC's assessment. However, not all clinical trials have monitoring guidelines; but should they? In this talk, I will focus on adaptive design clinical trials with pre-specified monitoring guidelines and how the DMC's formulation of recommendations based on a risk-benefit assessment are impacted by two factors: (1) the types of boundaries pre-specified; and (2) the information to be conveyed to the sponsor. I will consider the following questions which pertain to the DMC's efforts of protecting patients' interests and preserving integrity of monitored trials: Are these 'hard' or 'soft' boundaries (i.e. truly guidelines)? What information should be conveyed as part of the DMC's recommendations to the appropriate sponsor representative(s)? How might these boundaries differ in the setting of non-inferiority trials? What are any additional considerations for program-wide DMCs (single DMC monitoring multiple trials for the same investigational product)?

Contributors

A06

STRATEGIES TO IMPROVE PARTICIPANT RECRUITMENT TO RANDOMISED CONTROLLED TRIALS: A SYSTEMATIC REVIEW OF NON-RANDOMISED EVALUATIONS

HEIDI GARDNER

UNIVERSITY OF ABERDEEN

Background:

Randomised controlled trials (RCTs) guard against selection bias, offering the fairest way of evaluating healthcare interventions. Recruitment to trials can be difficult, and poor recruitment can lead to time and budget extensions. Many teams implement recruitment interventions as the trial is running in an attempt to improve poor recruitment. A substantial literature describing non-randomised evaluations of these interventions exists but has been rejected on grounds of anticipated poor methodological quality. However, systematic evaluation of this substantial body of work may provide useful information if interventions are similar enough for pooling and GRADE assessment supports higher levels of certainty than might be expected for a single study.

Methods:

The primary outcome for the review is the number of individuals, centres or physicians recruited. The methodological quality of studies was assessed using the Cochrane risk of bias tool for non-randomised studies (ROBINS-I).

We searched the Cochrane Methodology Register, MEDLINE, EMBASE, CINAHL and PsycINFO to identify non-randomised studies including a comparison of two or more interventions to improve recruitment. Two reviewers assessed abstracts and full texts for inclusion, disagreements were resolved with a third reviewer. We extracted data on the host trial, recruitment methods, design of the recruitment study, participant characteristics and setting, as well as study characteristics related to risk of bias.

Results:

We screened 9,642 abstracts, of which 223 full text articles were selected for assessment. 99 articles (102 studies) were eligible for inclusion; of those, 10 were at critical risk of bias and excluded in accordance with ROBINS-I guidance.

The final 92 included studies were all at serious risk of bias; 20 studies were included in a narrative analysis because they were at serious risk of bias in the 'confounding' domain only. All other studies were at serious or moderate risk of bias in two or more domains. Interventions spanned 7 categories: Use of networks/databases; face-to-face initiatives; language adaptations; postal invitations; recruiter awareness; recruitee awareness; randomisation methods. There was significant heterogeneity in interventions even within categories, making pooling studies unfeasible. Postal invitations were most frequently evaluated; appearing in 9 of the 20 studies, with a total of 894 participants. The most useful studies assessed interventions within one intervention category; e.g. language adaptations (English vs Spanish advertising recruited 563 and 2,012

participants respectively), and face-to-face recruitment initiatives (community institutions, academic institutions, and veterans' health administration sites recruited 1188, 1181, and 339 participants respectively).

Conclusion

Despite the effort and resources given to running and reporting non-randomised evaluations of recruitment interventions, their utility for trial decision-making is limited even when collated. We highlight the need for improved reporting; many of the included studies have omitted key pieces of information; specific details regarding interventions are often lacking, and limit use of the results generated. Non-randomised evaluations are unlikely to disappear; to improve their value we recommend clear distinctions between interventions, and detailed reporting using the TIDieR checklist.

This work is part of the Trial Forge initiative to improve trial efficiency.

Contributors

Loai Albarqouni
Bond University, Centre for Research in Evidence-Based Practice

Polly Black
NHS Lothian

Gordon Fernie
Centre for Healthcare Randomised Trials (CHaRT), University of Aberdeen

Katie Gillies
University of Aberdeen

Kirsty Loudon
University of Stirling

A07

CHARACTERISTICS AND TRENDS OF INDUSTRY SPONSORED CLINICAL TRIALS WITH A DMC – INSIGHTS FROM CLINICALTRIALS.GOV

KEHAO ZHU

AXIO RESEARCH

INTRODUCTION:

Data monitoring committees (DMCs) assess risks and benefits of participants and ensure the scientific integrity in ongoing clinical trials. While certain government sponsored clinical trials are required to appoint a DMC, for the industry sponsored trials, it is the sponsor's decision on the use of DMCs. FDA's guidance document, Establishment and Operation of Clinical Trial Data Monitoring Committees, discusses determining need for a DMC; e.g. DMCs are recommended for large trials that will compare rates of mortality. In this project, we analyzed the trials registered in ClinicalTrials.gov (CT.gov) to examine the characteristics and trends of DMC use in pharmaceutical industry sponsored clinical trials.

METHOD:

We accessed the CT.gov database and retrieved relevant information of each registered trial (e.g. whether a DMC is appointed) via the aggregate analysis of ClinicalTrials.gov (AACT), which is a part of Clinical Trials Transformation Initiative (CTTI). We identified the pharmaceutical industry sponsored trials with a start date between 2008 and 2017. We analyzed the study design factors associated with the decision of DMC use and the temporal trends of DMC use in the industry.

RESULTS:

We identified 36,755 pharmaceutical industry sponsored clinical trials started between 2008 and 2017. Of these trials, 9,869 (27%) trials were reported that a DMC was appointed, 21,692 (59%) trials were reported that a DMC was not appointed, and the remaining 14% of trials did not report this information. The rate of missing data in this DMC-use variable shrank from 19% in 2008 to 11% in 2017. Among the trials with this non-missing DMC-use variable, the percentage of DMC use increased from 22% in 2008 to 33% in 2017. Phase III trials have the highest rates of DMC use among the four major study phases, and these Phase III trials have the most rapid increasing trend in DMC use in the past 10 years with the percentage having a DMC increasing from 35% in 2008 to 55% in 2017. The industry sponsors are more likely to appoint a DMC to clinical trials if the trials are multicenter and have larger enrollment, longer duration, treatment arms masked, a survival endpoint as one of designed outcome measures and treatment as the primary purpose, although we noted that some studies with these design characteristics were reported that a DMC was not appointed. We will visualize these results of DMC use over time and the factors involved in DMC use for the presentation.

CONCLUSION:

In general, the pharmaceutical industry sponsors follow the FDA's guidance document on the necessity of establishing a DMC, and the use of DMCs is growing in the industry. The CT.gov database provided a unique source to quantify the characteristics and trends of DMC use in the industry, which provide a reference for the pharmaceutical industry sponsors. Lastly, we encourage the sponsors to accurately include information on whether a DMC is appointed, although it is not currently required by CT.gov.

Contributors

David Kerr
Axio Research

A08

A TARGETED APPROACH TO MEDICAL AND SAFETY REVIEWS

ERIK DOFFAGNE

CLUEPOINTS

Medical and safety monitoring is performed by reviewing a set of clinical parameters of interest, patient by patient. The current practice is to perform the review using paper based listings or visualization tools that support a manual approach.

This presentation will introduce a targeted approach relying on automated statistical methods to highlight atypical patients and clinical parameters. The session will demonstrate how patient profiles can provide detailed insight into patient experiences and offer a targeted approach by identifying anomalies in data and ranking patients by their relative degree of risk.

It will be shown that the approach helps to enhance overall quality management, guiding study teams to quickly and effectively characterize risk signals and enable centralized medical and safety reviews.

Learning Objectives

Participants will learn:

- How a targeted approach to medical review can be implemented.
- How to highlight atypical patients and clinical parameters.
- How to optimize the overall quality management by focusing on areas at risk.

Contributors

A09

RECRUITING MEN AGED 50-74 YEARS TO A LARGE DIABETES PREVENTION TRIAL: THE EFFECTIVENESS OF TRADITIONAL VS ONLINE PROMOTIONAL STRATEGIES

KAREN BRACKEN

UNIVERSITY OF SYDNEY

Introduction

Recruitment to clinical trials of disease prevention can be difficult and expensive. Previous reports have described recruiting participants using mass mail outs, media coverage, advertising, posters in the community and presentations at community and professional functions. More recently on-line strategies such as Google and Facebook advertising have been used. Much of the evidence for the efficacy of this online approach comes from studies recruiting young people or where the trial intervention is itself internet-based. There is limited evidence of the effectiveness of online strategies in recruiting older people to randomised controlled trials.

This study aims to compare the cost and effectiveness of traditional and online promotional strategies employed to recruit men aged 50-74 years to a study of testosterone for the prevention of type 2 diabetes (T4DM).

The T4DM study

T4DM is an Australian, publicly-funded, multicentre, double-blind, randomised, placebo-controlled trial investigating the efficacy of testosterone treatment, against the background of a lifestyle intervention, to normalise glucose tolerance in men aged 50-74 years with pre-diabetes or newly diagnosed diabetes and testosterone ?14 nmol/L. ANZCTR registration: ACTRN12612000287831

Methods

Initially study recruitment efforts focused on media news coverage and community-based promotions (posters, attendance at events). As recruitment progressed radio advertising, Facebook advertising, Google AdWords, inbound links (links to the study website on the websites of other organisations), postal and email newsletter mentions and mass mail outs were also introduced.

The effectiveness of each promotional strategy was assessed by measuring the potential reach (number of people approached, screened and randomised) and direct cost. Staff effort was also considered. Potential reach was given the highest priority as we estimated that 20,000 men would need to be screened to recruit our target sample size of 1000.

Results

In the four years of trial recruitment (Jan 2013-Feb 2017), 19,022 men were screened and 1007 men randomised. Most men were referred to the study by radio advertising (41%), media news coverage (25%) and mail outs (17%). Online strategies combined (Facebook and Google advertising, inbound links, email newsletters) accounted for only 5% of participant enrolments.

On average, promotional activities cost AU$567 per participant randomised. Specific strategies ranged from no direct cost (media coverage, community events, unpaid newsletters and email newsletters, inbound links) to AU$1,800 (newspaper advertising). Of the paid strategies, Facebook advertising (AU$590 per participant randomised) and mass mail out (AU$733 per participant) were comparable in cost while radio advertising (AU$1010 per participant) and newspaper advertising (AU$1800 per participant) were more expensive.

Conclusions

Online recruitment strategies had insufficient reach to replace traditional recruitment methods in this large disease prevention trial in men aged 50-74 years.

Contributors

Wendy Hague
University of Sydney

Anthony Keech
University of Sydney

Gary Wittert
University of Adelaide

A10

SURVIVAL BY HISPANIC ETHNICITY AMONG CANCER PATIENTS PARTICIPATING IN SWOG CLINICAL TRIALS

ANNA MOSELEY

SWOG STATISTICAL CENTER, FRED HUTCHINSON CANCER RESEARCH CENTER

SURVIVAL BY HISPANIC ETHNICITY AMONG CANCER PATIENTS PARTICIPATING IN SWOG CLINICAL TRIALS

Mariana Chavez-MacGregor, Joseph M. Unger, Anna Moseley, Scott Ramsey, Dawn L. Hershman

Background:

Racial disparities in cancer outcomes have been described. It is still unclear if patients of Hispanic ethnicity have better or worse survival outcomes. In this study we evaluated whether Hispanic participants in SWOG clinical trials had different survival outcomes compared to non-Hispanics.

Methods:

Adult patients registered in SWOG phase II/III clinical trials between 1986 and 2012 were analyzed. Studies of similar histology and stage were combined. Within each analysis, Kaplan-Meier survival curves were generated to examine differences in outcome by ethnicity. Multivariate Cox regression was used to estimate the association of ethnicity and survival outcomes, controlling for major disease-specific prognostic factors and demographic variables plus area-level income and education to account for socioeconomic status.

Results:

29,338 patients registered to 38 trials were included; 5% of them were Hispanic. Hispanics were more likely to be younger and from lower income and education areas (all $p<0.05$). No differences in survival were observed across tumor types except in the advanced-stage prostate cancer group, where we observed an association between Hispanic ethnicity and worse overall survival (HR=1.40, p=0.006), progression-free survival (HR=1.36, p=0.007), and cancer specific survival (HR=1.42, p=0.013). After adjusting for multiple comparisons, no differences in outcomes were seen.

Conclusions:

Hispanic patients participating in SWOG trials receiving uniform treatment and follow-up had similar survival outcomes compared to non-Hispanic patients, with the single exception of the advanced stage prostate cancer group.

Impact:

Despite a large literature describing racial disparities in cancer, our study shows that Hispanic patients receiving uniform treatment and follow-up have similar outcomes compared to non-Hispanics.

Contributors

Mariana Chavez-MacGregor
The University of Texas MD Anderson Cancer Center

Joseph Unger
SWOG Statistical Center, Fred Hutchinson Cancer Research Center

Scott Ramsey
SWOG Statistical Center, Fred Hutchinson Cancer Research Center

Dawn Hershman
Columbia University

A11

COMPARISON OF MASS MAILINGS TO FACEBOOK ADVERTISING FOR RECRUITMENT OF MIDLIFE WOMEN WITH BOTHERSOME VAGINAL SYMPTOMS

KATHERINE GUTHRIE

FRED HUTCHINSON CANCER RESEARCH CENTER

Background: The MsFLASH (Menopause Strategies: Finding Lasting Answers for Symptoms and Health) Network recruited healthy volunteers into four randomized clinical trials, ranging in sample size from 100 to 350 women, primarily through mass mailings. For a 5th trial, in addition to mass mailings we piloted a social media recruitment approach to evaluate utility and cost relative to mailings. This trial tested two interventions for postmenopausal vulvovaginal symptoms (dryness, itching, irritation, pain with sex), and thus required a higher level of sensitivity to privacy concerns compared to our previous studies of hot flashes and insomnia.

Objectives: To evaluate our ability to recruit women for the Vaginal Health Trial through Facebook advertising. Also, to compare the cost per randomized woman recruited through mass mailings compared to Facebook.

Methods: Design of recruitment materials was informed by focus group feedback. For mail recruitment, we purchased commercially available mail lists targeted by age (55-70) and zip codes within a 20-mile radius of the Minneapolis and Seattle clinical sites. A commercial mail house sent 10-20,000 invitation letters by bulk mail every 2 weeks. Facebook ads were targeted to women aged 50-65+ within 20 miles of each city center, with ads not running continuously but posted to and taken down from the website as needed to control flow to the clinics. The Fred Hutchinson Cancer Research Center's Communications & Marketing Department supported recruitment letter design, MsFLASH website creation, Facebook ad design, and managed the Facebook ad campaign. We compared costs and eligibility rates at each step of recruitment and follow-up using descriptive statistics.

Results: 277,000 letters were sent in the first 5 months of recruitment (4/11-9/1/16), resulting in 278 women randomized at a cost of $98,682. Mailing costs were approximately equal across sites. Response to mailings was greater than expected, thus fewer spaces than planned were available for women recruited via social media. Facebook ads ran in Minneapolis for 28 days total during 9/27-12/9/16, and in Seattle for 15 days total during 10/19-12/9/16. Recruiting by social media, 25 women were enrolled at a cost of $14,813. Across all screening steps, women recruited by mail were more likely to be eligible and thus to be randomized, than those recruited by Facebook (13% vs. 5% of respondents were randomized, respectively). Overall, the cost of recruitment by mail was lower ($356) than that via Facebook ($593) per randomized participant. Notably, women in Seattle responded more frequently to Facebook ads compared to women in Minneapolis (15.0 vs. 8.4 responses per ad day), thus the cost of recruiting by Facebook in Seattle was only $122 per woman randomized versus $1,174 per woman in Minneapolis.

Conclusions: Recruitment to a clinical trial testing interventions for menopausal vaginal symptoms can be achieved through social media advertising, although costs per randomized participant are

not necessarily less than with mass mailings. Variability in observed cost may reflect the small sample sizes, as well as other confounding factors such as better recognition of the Fred Hutchinson Cancer Research Center logo on Facebook ads in the Seattle area.

Contributors

Bette Caan
Kaiser Permanente of Northern California

Susan Diem
University of Minnesota

Kristine Ensrud
University of Minnesota

Sheri Greaves
Fred Hutchinson Cancer Research Center

Andrea LaCroix
University of California at San Diego

A12

IMPROVING THE EVIDENCE BASE FOR TRIAL PROCESS DECISION-MAKING: STUDIES WITHIN A TRIAL (SWAT) AND WHY WE SHOULD BE COORDINATING AND DOING MORE OF THEM

SHAUN TREWEEK

UNIVERSITY OF ABERDEEN

The problem

The Director of the UK's main publicly-funded trial program, the National Institute for Health Research Health Technology Assessment program describes clinical trials as the backbone of primary research informing clinical practice in the National Health Service. However, the evidence available to trial teams to inform their trial process decisions–e.g. how to recruit participants, how to retain them, what sort of site visits are required–is close to zero. The literature is characterised by single evaluations, often poorly done, of interventions with no clear design rationale. A strong body of evidence that trial teams can trust is yet to emerge from this.

Studies Within A Trial–SWATs

One way to fill gaps in trial methods evidence is to run Studies Within A Trial, or SWATs. SWATs are a self-contained research study that has been embedded within a host trial with the aim of evaluating or exploring alternative ways of delivering or organising a particular trial process. They can provide information of immediate relevance to decisions in the host trial but can also provide evidence for synthesis in systematic reviews that will improve the conduct of future trials.

This talk will introduce SWATs, explain why they are important and how to design them as well as providing some practical 'top tips' from the experience of doing SWAT in the UK and Ireland. Experience shows that SWATs need not be expensive, are ethically acceptable (in the UK and Ireland some types have not needed formal approval by an Ethics Committee) and analysis is often simple. There are challenges though, which will be discussed. They work best when done as part of a coordinated and collaborative initiative to reduce uncertainty around high priority trial process alternatives.

Two concrete examples of collaborative SWATs will be presented, including one evaluating the effect on recruitment of a structured way of developing participant information materials. This was tested in four trials in three years, involved nearly 8000 people and gave a clear result: little or no effect on recruitment (absolute increase = 1% (95% CI = -1% to 2%); GRADE assessment: high certainty). The recent update of the Cochrane systematic review of interventions to improve recruitment (which ST leads) finds that the global literature supports this level of certainty for just two other interventions. The evidence behind the other 69 interventions in that review is far less certain and, even for the two other high certainty interventions, little has changed since 2004.

A SWAT protocol together with coordinated, collaborative evaluation can take evidence from a standing start to evidence-informed decision making in a few years. The current approach to finding ways to improve trial processes takes decades.

Conclusion

SWATs are a key tool in the trial methods armoury. Coordinated SWATs can produce high certainty evidence for trial process questions that would otherwise take decades to answer. They are generally cheap, ethically acceptable, simple to analyse and tailor-made for synthesis. We should be doing more of them.

Contributors

Declan Devane
National University of Ireland Galway

Seonaidh Cotton
University of Aberdeen

Adel El Feky
University of Aberdeen

Heidi Gardner
University of Aberdeen

Katie Gillies
University of Aberdeen

A13

DEVELOPMENT AND IMPLEMENTATION OF A WEB-BASED SAE ADJUDICATION SYSTEM.

KALEAB ABEBE

UNIVERSITY OF PITTSBURGH

Unbiased ascertainment of primary and secondary outcomes is critical to the internal validity of any randomized controlled trial. This is especially important in unblinded studies or studies where the outcomes are subjective in nature. The use of adjudication or clinical events committees is one mechanism to remove potential bias. Typically, this involves a committee comprised of members independent of the trial or a subgroup of study investigators. For each outcome or event in question, a primary and secondary (and sometimes, tertiary) reviewer provides their overall assessment (i.e. PKD-related hospitalization, SCD-related pain event, etc), which may differ from the initial assessment. Furthermore, adjudicators may disagree with one another, but ultimately have to arrive at consensus.

Despite the straightforward conceptual model of an adjudication process, the implementation of such a process in a clinical trial is easier said than done. Many times, the planning for adjudication starts well after the initiation of a clinical trial; in some cases, during the data cleaning process. This delay can result in incomplete information if any analyses are done using the non-adjudicated data (i.e. interim analyses for DSMB). If the adjudication is put off until the end of the trial, recall bias or personnel changes can result in missing information that could have been used to aid adjudication (i.e. missing hospital discharge summaries). As a result, it is imperative to have a well-designed adjudication process that is integrated within the electronic data capture system at the outset of a study.

Two examples are the recently completed HALT-PKD and ongoing STERIO-SCD trials. HALT-PKD was comprised of two concurrent double-blind, placebo-controlled clinical trials which investigated the impact of low versus standard blood pressure control (Study A) and dual versus single blockade of the renin-angiotensin-aldosterone system (Study A & B) on the progression of autosomal dominant polycystic kidney disease (ADPKD). The primary outcome for Study A was annual percent change in total kidney volume, and the primary composite endpoint for Study B was time to death, end-stage renal disease, or a 50% reduction in estimated glomerular filtration rate (eGFR) from baseline. For both studies, one of the secondary outcomes included PKD-related hospitalizations. The Safety, Tolerability, and Efficacy of Riociguat in Patients with Sickle Cell Disease (STERIOD-SCD) trial is a phase IIb, randomized, double-blind, placebo-controlled study to evaluate the safety and tolerability of 12-weeks of treatment with riociguat versus placebo in high-risk patients with sickle cell disease (SCD). Safety and tolerability monitoring include physical exams, adverse effect assessment, and laboratory evaluations. One of the secondary outcomes is SCD-related serious adverse events.

In this talk, we will describe the development and implementation of a web-based adjudication portal within the STERIO-SCD as well as outline the challenges encountered. Additionally, we will contrast this with the HALT-PKD trials, where the adjudication process was not implemented until much later in the study. Our hope is that this results in a “lessons learned” for clinical trialists.

Contributors

Sharon Stover
University of Pittsburgh

Jason Kojtek
University of Pittsburgh

Susan Spillane
University of Pittsburgh

A14

PRAGMATIC TRIAL MANAGEMENT: EXPLORING EVIDENCE FOR TRIAL PLANNING MILESTONES

PAULA DARBY LIPMAN

WESTAT

Introduction

Improving efficiencies across all phases and types of clinical research is crucial to accelerating translation of findings into practice leading to better delivery of effective, patient-centered care. Dissemination of lessons learned regarding effective trial planning and conduct can establish more reliable metrics for early identification of at-risk trials and also contribute to management guidance and standardization. This presentation describes outcomes of a program that monitored pragmatic clinical trials (PCTs) by working closely with investigators to identify and track achievement of trial planning milestones.

Background

The NIH Pragmatic Trials Collaborative Project was initiated in 2014 to support scientifically diverse low-cost, randomized-controlled clinical intervention trials. Assessing achievement of planning milestones and metrics, facilitated by an independent evaluation and coordinating unit (Westat), formed the evidence base to substantiate administrative review and approval for continued funding.

Methods

Six principal investigators (PIs) and members of their project teams participated in monthly conference calls during the first year of funding to discuss progress meeting milestones agreed upon in collaboration with NIH Program Officers (POs). Milestones (and associated metrics for which delivery of documented evidence was required) were organized by category (collaborations, materials and methods, clearances, study population, recruitment, resources and patient information management), with those falling outside these categories classified as trial-specific. Interviews were conducted with PIs, team members, and NIH POs to discuss the impact of the monitoring approach.

Results

For the five of six trials transitioning to the 4-year trial phase, the number of milestones ranged from 6-15. Metrics ranged from 15-33, for a total of 121 deliverables. One third of the metrics (42, 35%) were trial specific, with the remainder falling in one of the standard categories. During monthly calls there was considerable discussion around applicability of terms and concepts, contextual challenges such as stakeholder buy-in and timeline constraints, and appropriate deliverables especially for more technical milestones such as databases. Members of the study teams reported that the monitoring was onerous at times but improved or complemented their

management strategies, and POs found the extensive documentation submitted by Westat provided sufficient evidence to assess trial readiness during the administrative review.

Conclusions

Implementing systematic approaches to identify and track milestones can strengthen the evidence base regarding time and effort required to efficiently conduct and manage clinical trials. Although each awardee in this small project was required to provide evidence of completion of metrics, there was considerable variability in number and type required. Investigators were unaccustomed to producing documentation of performance metrics and in some cases it was challenging to determine what objective evidence to provide. Efforts to standardize expectations regarding milestones that mark a significant change or stage in trial development may provide guidance for determining what evidence of likely success "looks like." A typology or framework with clear specification of metrics is especially critical for transparency, particularly when funding decisions are contingent on both merit and feasibility.

Contributors

Leanora Dluzak
Westat

Catherine Stoney
National Heart, Lung, and Blood Institute (NHLBI), National Institutes of Health (NIH)

A15

THE STANDARD ERROR OF THE TREATMENT EFFECT IN STEPPED WEDGE TRIALS WHEN THE WITHIN-CLUSTER CORRELATION STRUCTURE IS MISSPECIFIED

JESSICA KASZA

MONASH UNIVERSITY

Introduction

Multiple-period cluster randomised trials, such as stepped wedge or cluster cross-over trials, are being conducted with increasing frequency. In the design and analysis of these trials it is usually assumed that the correlation between the outcomes of any pair of subjects is identical: the Hussey and Hughes within-cluster correlation structure, which encodes this assumption, is the most frequently assumed (1). However, more complex models that allow for correlations between pairs of subjects to decay over time have recently been suggested, and may be more appropriate for many studies and outcomes (2). Although the impact of these more complex within-cluster correlation structures on treatment effect estimator variances and required sample sizes is known, how estimated variance components are affected by within-cluster correlation structure misspecification is not yet understood. In addition, the impact of this misspecification on the standard error of the estimated treatment effect requires investigation.

Methods

We seek to determine under which circumstances incorrectly assuming equal correlations within a cluster (the Hussey and Hughes model) leads to poor coverage of the confidence interval of the treatment effect estimate. Assuming balanced data, we consider ANOVA estimators of variance components, and examine the impact of within-cluster misspecification at the planning stage of a trial on required sample sizes, and at the analysis stage on the treatment effect standard error.

Results

Incorrectly assuming the Hussey and Hughes model when within-cluster correlations decay over time can lead to over- or under-estimation of the standard errors of treatment effect estimates: the direction of the mis-estimation depends on the trial design and the true correlation structure. For many common stepped-wedge designs, incorrectly assuming the Hussey and Hughes model will lead to underestimation of the standard error of the treatment effect; while for the cluster cross-over design, this incorrect specification will lead to overestimation of standard errors. The impact of within-cluster correlation structure misspecification on treatment effect estimate standard errors is minimal for multiple-period parallel designs.

Conclusions

In the analysis of data from cluster randomised trials that incorporate treatment switches, such as the cross-over and stepped wedge designs, misspecification of the within-cluster correlation structure can have a large impact on estimated variance components. This impacts both sample size calculation and treatment effect standard error estimation. Failure to correctly model within-

cluster correlation structure can lead to incorrect conclusions regarding estimated treatment effects for stepped wedge and cluster crossover trials.

References

1. Hussey MA, Hughes JP. Design and analysis of stepped wedge cluster randomized trials. Contemporary Clinical Trials 2007;28(2):182-91.

2. Kasza J, Hemming K, Hooper R, et al. Impact of non-uniform correlation structure on sample size and power in multiple-period cluster randomised trials. Statistical Methods in Medical Research 2017; DOI: 10.1177/0962280217734981.

Contributors

Andrew Forbes
Monash University

A16

THE WIN RATIO APPROACH TO THE ANALYSIS OF A COMPOSITE OUTCOME IN THE MULTIPLE RISK FACTOR INTERVENTION TRIAL

ALES KOTALIK

UNIVERSITY OF MINNESOTA

Background:

The Multiple Risk Factor Intervention Trial (MRFIT) was a randomized trial of a multifactor intervention on coronary heart disease (CHD) in men. The trial did not yield a significant difference for the primary endpoint of CHD death for the special intervention (SI) versus usual care (UC) groups (hazard ratio (HR) =0.93; 95% CI: 0.72–1.20). However, a major limitation of the trial was that it was under-powered, as fewer deaths occurred than was anticipated.

Methods:

To address the loss of power, a composite cardiovascular disease (CVD) outcome including a priori defined components was defined and reported using a traditional first-event analysis, which only considers time until the first occurrence of any of the components (and therefore treats all outcomes as if they had equal importance). The CVD composite (sorted by severity measured as 20-year risk of post-trial CVD death) consists of: CVD death, renal impairment, coronary heart failure, myocardial infarction (MI) by serial ECG, MI by hospital records, surgery for coronary artery disease, and stroke.

Since the components vary in severity, we re-analyze the data using the matched win ratio approach proposed by Pocock with matching on the Pooled Equation risk score and duration of follow-up. This approach first pairs the participants in the SI and UC groups and then simply notes which of the pair experienced the highest severity event first. If it is not possible to untie the pair due to censoring, the comparison moves on to the second order event and so on. The resulting win ratio is defined as number of wins for the SI group divided by number of wins for the UC group. The inverse of the win ratio (the loss ratio) is then directly comparable to previously reported HR (SI/UC).

We also apply an alternative win ratio approach proposed by Oakes, which uses the proportional hazard model and modifies the risk set for each of the events by censoring participants if they experienced a greater severity event. For this reason, a single participant can be included in multiple risk sets. The proportional hazards model of Wei, Lin and Weissfeld (WLW) is then applied to account for the within-subject correlation.

Results

The resulting HR and confidence intervals using the aforementioned approaches (primary endpoint analysis, first event CVD composite, Pocock and Oakes win ratios) are summarized in Table1, along with number of events used. The point estimates are similar across methods for summarizing the CVD composite, though the number of events varies widely.

Summary

A borderline significant difference between the SI and UC groups in MRFIT is found using the win ratio approach. This approach takes into account the varying importance of events in the composite, an issue often overlooked in the analysis of composite outcomes. The result is consistent with previously reported analyses of the MRFIT trial, utilizing a first event analysis for the CVD composite.

Contributors

Anne Eaton
University of Minnesota

Qinshu Lian
University of Minnesota

Carlos Serrano
University of Minnesota

John Connett
University of Minnesota

Jim Neaton
University of Minnesota

A17

VARIATIONS IN DESIGN OF THE STEPPED WEDGE CLUSTER RANDOMIZED TRIAL: ILLUSTRATION OF STATISTICAL CONCEPTS USING THE PATIENT-CENTERED CARE TRANSITIONS IN HEART FAILURE (PACT-HF) TRIAL

SHUN FU LEE

POPULATION HEALTH RESEARCH INSTITUTE

The Patient Centered Care Transitions in Heart Failure (PACT-HF) trial is a stepped wedge (SW) cluster randomized trial that evaluates the effectiveness of a complex health service intervention among patients hospitalized for HF across several large hospitals. A SW design has been selected so that all the participating clusters (hospitals) will eventually receive the intervention, which is expected to be of benefit and unlikely to cause harm to patients. The sequential implementation of the intervention allows for the lead time required for successful implementation of services. Another benefit of the SW design is that compared to a parallel cluster randomized trial, fewer clusters are required to statistically power the study.

The aim of this study is to explore the statistical power implications for variations of the SWD that may be required to mitigate logistic challenges to implementation of complex health services in real-world setting. The PACT-HF trial involves 10 hospitals and 11 time periods. The primary endpoint is the time to composite readmissions/emergency department visits/death at 3 months. Secondary outcomes, including patient-reported measures, will be analyzed in nested study involving 8 hospitals and 9 time periods. Thus, PACT-HF is a complete SW trial with a nested study. Options for SW trial can also include block stratification of sites according to readiness, with clusters randomized to begin the intervention within an “early” or “late” block; an incomplete SW design with data collection occurring only during select time periods; and a hybrid trial design that combines parallel cluster with SW design. The statistical power considerations for the variations in SW trial design are assessed in this study.

By understanding the methodological and statistical implications of different SW designs, researchers can increase the likelihood of successfully completing their trial with adequate statistical power.

Funded by Canadian Institutes of Health Research (CIHR), Ontario’s Ministry of Health and Long Term Care. PACT-HF NCT02112227

Contributors

Rudy Unni
Michael G. Degroote School of Medicine, McMaster University

Harriette Van
Department of Medicine, McMaster University

A18

EXPLORING THE HAWTHORNE EFFECT USING A BALANCED INCOMPLETE BLOCK DESIGN IN THE ASPIRE CLUSTER RANDOMISED CONTROLLED TRIALS

MICHELLE COLLINSON

UNIVERSITY OF LEEDS, CLINICAL TRIALS RESEARCH UNIT

Background:

The Hawthorne effect is a non-specific treatment effect: an alteration in behaviour as a response to the interest or attention received through observation and assessment leading to an overestimate of the effectiveness of an intervention [1, 2]. If the Hawthorne effect is unbalanced across trial arms, resulting treatment estimates may be biased [2].

Action to Support Practices Implementing Research Evidence (ASPIRE) is a UK NIHR-funded programme aiming to develop and evaluate interventions to promote adherence to evidence-based quality indicators in general practice. Multifaceted implementation packages were adapted to target four indicators and evaluated using anonymised, routinely collected electronic health records in two parallel cluster randomised controlled trials (cRCTs) in West Yorkshire general practices.

Methods:

Balanced incomplete block designs, with each trial having two arms, were intentionally chosen to equalise Hawthorne effects whilst maximising power and efficiency [3, 4]. Each trial evaluated the effect of adapted implementation packages on adherence to two of four indicators: diabetes control and risky prescribing in trial 1; blood pressure control and anticoagulation in atrial fibrillation in trial 2. Within trials, implementation packages were assumed to be independent with respect to their effect on the indicators, thus practices randomised to the implementation package for one indicator acted as control practices for the other implementation package and vice versa.

A fifth, non-intervention control group was included to enable a secondary analysis testing for evidence of Hawthorne effects. General practices allocated to this control group received none of the ASPIRE adapted implementation packages. According to the theory, if a Hawthorne effect is present the non-random aspect of the differences in the intervention effects in the primary analysis can be attributed to the fact that practices were aware of being observed (resulting in improvements to adherence to indicators) and is not attributable to the intervention. Further, we expect the intervention effect in the primary analysis will be smaller than in the secondary analysis utilising the non-intervention control practices.

Results:

ASPIRE recruited 178 general practices using an opt-out approach; 80 randomised to trial 1; 64 randomised to trial 2; with 34 randomised to the non-intervention control group. The implementation package reduced risky prescribing (OR=0.82, 97.5% CI (0.67 – 0.99)) but had no statistically significant effect on the other primary endpoints. We will present an analysis exploring

the Hawthorne effect in both trials, using the non-intervention controls and discuss the implications for future implementation trials.

Conclusions:

Using a balanced incomplete block design and including randomised non-intervention controls could inform the design and analysis of future RCTs, particularly those utilising routinely collected data in implementation research.

References:

1. Sedgwick P, Greenwood N. Understanding the Hawthorne effect. BMJ. 2015:351:h4672.

2. Grimshaw J et al. Experimental and quasi-experimental designs for evaluating guideline implementation strategies. Family Practice. 2000 Feb;17 Suppl 1:S11-6.

3. Eccles M et al. Research designs for studies evaluating the effectiveness of change and improvement strategies. Qual Saf Health Care. 2003;12(1):47–52. doi:10.1136/qhc.12.1.47.

4. Grimshaw JM, et al. Effectiveness and efficiency of guideline dissemination and implementation strategies. Health Technol Assess. 2004;8(6):iii-iv, 1-72.

Contributors

Thomas Willis
University of Leeds, Leeds Institute of Health Sciences

Robbie Foy
University of Leeds, Leeds Institute of Health Sciences

Liz Glidewell
University of Leeds, Leeds Institute of Health Sciences

Suzanne Hartley
University of Leeds, Clinical Trials Research Unit

Paul Carder
NHS Bradford Districts Clinical Commissioning Group

A19

METHODS, SUCCESSES, AND LESSONS LEARNED FROM CENTRALIZED COGNITIVE ASSESSMENT IN A MULTICENTER CLINICAL TRIAL OF CAROTID REVASCULARIZATION AND MEDICAL MANAGEMENT FOR ASYMPTOMATIC CAROTID STENOSIS (CREST-2)

VIRGINIA HOWARD

UNIVERSITY OF ALABAMA AT BIRMINGHAM

Objective: To describe the methodology, successes and lessons learned related to central administration of a battery of cognitive tests in a multicenter clinical trial.

Background: The Carotid Revascularization and Medical Management for Asymptomatic Carotid Stenosis Trial (CREST-2) is two trials assessing: 1) treatment differences between intensive medical management (IMM) alone compared to carotid endarterectomy plus IMM, and 2) treatment differences between IMM alone compared to carotid stenting plus IMM in the primary prevention of stroke in patients with high-grade asymptomatic carotid stenosis. (US NIH clinicaltrials.gov NCT02089217) A secondary outcome is change in cognitive function. Approximately 120 clinical sites across North America are participating. Randomization began December 2014 and is ongoing for a targeted sample size of 1240 in each trial.

Methods: Participants undergo cognitive assessment prior to revascularization or starting the IMM (baseline), and at 44-day, 12-month and annually thereafter to 48 months. The neurocognitive battery is administered centrally via telephone by rigorously trained interviewers from the Survey Research Unit (SRU), University of Alabama at Birmingham. This assessment includes five tests comprising four domains of cognitive function: Learning (CERAD Word List Learning Test), memory (CERAD Word List Delayed Recall), executive function/processing speed (animal naming and letter fluency), and attention/working memory (digit span). Prior to the participant's in-person visit (or telephone follow-up in special cases), the site coordinator makes a reservation for the cognitive assessment using the web-based reservation program in the CREST-2 Electronic Data Entry System (eDES). The reservation holds a place to guarantee the availability of a trained interviewer. Training and guidelines for clinic personnel are provided such as having a quiet and private area in the clinic clear of paper, pencils, pens; plan for when in the study the cognitive assessment will take place, e.g., early in day and early in visit is best, allowing for a full 30 minutes. In rare cases when a cognitive assessment cannot be made in the clinic, the site coordinator still uses the eDES reservation system and provides similar guidelines to the participant.

Results: As of October 16, 2017, 114 sites have randomized at least one patient. Of all expected cognitive assessments, 86% have been completed with a high percentage at baseline (91%), and a somewhat lower percentage at 44-days (84%), 12-months (79%), and 24-months (77%, 74/96). This lower rate at follow-up is a product of some participants declining to participate in the cognitive assessment and poor performance at a relatively small number of sites. Repeated reminders to use the reservation system have been required. Feedback relative to participant burden in proximity of cognitive assessments (baseline and 44-days) and too close to revascularization in half the patients has prompted a protocol amendment to remove the 44-day

assessment. Analyses of baseline cognitive assessment of the first 207 participants support the sensitivity of this battery to detect cognitive function.

Conclusions: Cognitive assessment that includes five tests of cognitive function can be administered centrally, with a high completion rate, providing a sensitive battery in a multicenter clinical trial to detect changes in cognitive function.

Contributors

Virginia Wadley
University of Alabama at Birmingham

Paul Wolff
University of Alabama at Birmingham

Yu Zhang
University of Alabama at Birmingham

David Knopman
Mayo Clinic

Brajesh Lal
University of Maryland School of Medicine

A20

INTERVENTIONS TO IMPROVE RETENTION TO CLINICAL TRIALS: A BEHAVIOURAL ANALYSIS OF EFFECTIVE INTERVENTIONS.

KATIE GILLIES

UNIVERSITY OF ABERDEEN

Background

Clinical trials often struggle to retain the number of participants required on which to make valid and reliable assessments about effectiveness of treatments. Several individual randomised comparisons of interventions to improve retention in trials have been shown to be effective. However, the effect across trials is inconsistent (Brueton et al., 2013).

Many of these retention interventions target participants' behaviour (e.g. returning questionnaires or attending a follow up visit).Although not designed as such, these interventions can be thought of as behaviour change interventions with, for example, returning a questionnaire being the behaviour targeted. By coding the constituent behaviour change components of effective retention interventions, we identified the interventions' potential 'active ingredients' responsible for improvements in retention and will highlight these in this presentation.

Methods

Studies reporting effective (defined by reports testing retention interventions that have been included in a meta-analysis and show evidence of improvements in retention) retention interventions were identified from existing meta-analyses in the literature. Published manuscripts, intervention and control group material, and associated paperwork provided in response to requests to authors were coded into their behaviour change techniques (BCTs) using the BCT taxonomy version 1. Two authors independently coded materials using a standardised coding manual and discussed any disagreements to reach consensus. Data on study characteristics including host trial context, timing, mode of delivery and dosage of retention intervention were recorded.

Findings

Two intervention types were identified as having evidence of improving retention in existing meta-analyses; monetary incentives and text message prompts, with both having their effect on the return of postal questionnaires. No effective intervention was identified to support attendance at face-to-face visits. None of the interventions identified explicitly stated a theoretical rationale for their development or implementation. The BCTs used in 'monetary incentive' interventions differed to the control group by use of the BCTs 10.2 material reward (behaviour), 10.8 incentive (outcome), or by 10.10 reward (outcome). Contrastingly, the BCTs identified in 'text message prompts' interventions were identical in both the control and intervention groups and differed only in terms of mode of delivery and dosing.

Conclusions

Attending a measurement visit or returning a questionnaire is a behaviour and trialists should be mindful of this when designing retention interventions. Our work in this area provides some of the first evidence of the impact of implicit use of BCTs in retention intervention and highlight their promise for future. This presentation will discuss the use of BCTs in interventions to improve retention in clinical trials, the evidence of their use to date and their potential for future inclusion and evaluation. The promise of theoretical perspectives to inform and improve aspects of trial process and delivery have begun to be realised. The potential benefits of this approach are likely to outweigh any additional implications and therefore could be considered a key component of trial design decisions.

Contributors

Tina Bennett
University of Aberdeen

Eilidh Duncan
University of Aberdeen

A21

A TWO-STAGE VERSION OF FISHER'S EXACT TEST FOR MULTI-ARM TRIALS WITH BINARY OUTCOME VARIABLES

MICHAEL GRAYLING

UNIVERSITY OF CAMBRIDGE

In small sample trials with binary outcome data, use of a normal approximation for hypothesis testing can lead to substantial inflation of the type-I error-rate. Consequently, exact statistical methods are required, and accordingly, much research has now been conducted to facilitate this. Recently, this included methodology for two-stage multi-arm trials utilising exact binomial tests [1]. These designs were demonstrated to carry substantial efficiency advantages over a fixed sample approach, but unfortunately suffered from strong conservatism.

An alternative means of small sample inference with binary data is Fisher's exact test. However, this method is limited to single-stage designs when there are more than two treatment arms [2]. Therefore, here, we propose a two-stage version of Fisher's exact test, with the potential to stop early to declare treatment efficacy or treatment futility, which is applicable to multi-arm trials.

We first define the probability of committing a familywise error after the first and second stage of such a trial, conditional on the employed stopping boundaries and the observed number of patient responses. We are then able to use the error spending approach to group sequential trial design, in combination with an efficient grid search, to determine optimised two-stage versions of Fisher's exact test.

For a motivating example based on a phase II oncology trial, we demonstrate that on average our approach is less conservative than the corresponding optimal exact binomial test designs. Moreover, our design is able to reduce the sample size required under the alternative hypothesis for this example by 11%, whilst simultaneously providing an increase in power.

In all, our two-stage Fisher's exact test may routinely be expected to provide more desirable efficiency gains than a sequential exact binomial testing procedure. This could greatly assist the design of phase II randomised trials, where large required sample sizes are considered particularly unfavourable.

[1] Jung S-H (2008) Randomized phase II trials with a prospective control. Stat Med 27:568-83.

[2] Jung S-H, Sargent DJ (2014) Randomized phase II cancer clinical trials. J Biopharm Stat 24:802-16.

Contributors

Adrian Mander
University of Cambridge

James Wason
University of Cambridge

A22

PHASE I TRIALS OF IMMUNOTHERAPIES AND MOLECULARLY TARGETED AGENTS: WHAT DESIGNS ARE USED, AND HOW WELL ARE THEY REPORTED?

GRAHAM WHEELER

UNIVERSITY COLLEGE LONDON

Background: Immunotherapies and Molecularly Targeted Agents (IMTAs) for treating cancers are designed to target tumour cells or tumour-specific pathways whilst managing patient exposure to toxicities. In phase I trials, identifying a recommended phase II dose is often based on controlling the risk of inducing severe treatment-related toxicity. Many IMTAs have different anticancer and toxicity profiles to cytotoxic chemotherapy, which presents a challenge from a dose-finding perspective. Treatments may provide benefit at non-toxic doses, and dose-efficacy relationships may not be monotonically increasing. Several novel model-based designs for dose-finding studies of IMTAs have been proposed in the statistical literature, but it is unclear how often they have been implemented in practice. We present a detailed review of recently published phase I trials of IMTAs and their designs, and compare our findings to trends observed in past reviews.

Methods: We searched PubMed for phase I oncology trials designed to estimate the recommended phase II dose for an IMTA administered either as monotherapy or as part of a combination therapy. Eligible trials were published between January 2015 and March 2017 and were assessed for the type of dose-escalation design employed, treatments tested, tumour histology, country of trial sponsor, and journal of publication.

Results: We found 322 trials, of which 176 were eligible for our review. One hundred and five trials (60%) were on IMTAs given as monotherapy, with 15 trials (9%) investigating combinations of at least one immunotherapy and one MTA. The majority of trials were sponsored by centres in the United States of America (103; 59%). The traditional 3+3 design was used in 114 trials (65%), while only nine trials (5%) used a model-based design; this is slightly lower than the results from a review of IMTA studies published between 2008 and 2014 (55/820 trials (6.7%), Fisher's exact test $p = 0.50$). The remaining studies either used other rule-based designs (36; 20%), or did not describe the design in enough detail to determine how treatments were escalated (17; 10%).

Conclusions: In the current review, the majority of phase I trials of IMTAs used rule-based approaches to identify recommended phase II doses. This is despite concerns raised that rule-based designs are inappropriate for phase I trials of targeted anticancer therapies, and that novel designs have been proposed explicitly for such studies. Furthermore, one in ten trials do not report the study design to an appropriate standard. Improved efforts are needed to consider more appropriate trial designs for phase I studies of IMTAs, and to report trial designs clearly and thoroughly in publications.

Additional work: Further to the current review, we present updated work to include published trials up to December 2017. We also investigate the number of patients and dose levels used, and discuss the dose-limiting toxicity and response rates observed to better describe the current landscape of phase I trials for IMTAs.

Contributors

Aikaterini Florou
Guy's and St. Thomas' NHS Foundation Trust

Hakim-Moulay Dehbi
University College London

A23

REACH OUT: RANDOMIZED CLINICAL TRIAL OF EMERGENCY DEPARTMENT-INITIATED HYPERTENSION BEHAVIORAL INTERVENTION CONNECTING MULTIPLE HEALTH SYSTEMS

MACKENZIE DOME

THE UNIVERSITY OF MICHIGAN

BACKGROUND

Hypertension, disproportionally impacts African Americans, and is the most important modifiable risk factor for cardiovascular disease, the leading cause of mortality in the United States. The Emergency Department (ED) represents a missed opportunity to identify and treat hypertension, within difficult-to-reach populations. Reach Out is a multicomponent, health theory based, mobile health behavioral intervention to reduce blood pressure among hypertensive patients evaluated in a safety net ED in Flint, Michigan. Utilizing an innovative clinical trial philosophy, the Multiphase Optimization Strategy (MOST), will allow Reach Out to efficiently determine optimal intervention components.

METHOD

Reach Out is a randomized, controlled, 2x2x2 factorial design clinical trial that consists of three components, each with two levels or 'dose'; healthy behavior text messaging; prompted home blood pressure self-monitoring; facilitated primary care provider appointment scheduling and transportation. Subjects will be identified in a safety-net ED and enrolled in the trial. Randomization will take place after a 3 week screening period to ensure the presence of hypertension outside of the ED.

Traditionally, randomized control trials are limited when evaluating multicomponent behavioral interventions resulting in uncertainty regarding the performance of each individual intervention component. As a result, it is unknown whether each component's effect size justifies the resources required to implement it. An alternative to traditional two-arm trials comparing multicomponent interventions to control is the MOST strategy, which involves factorial design, to determine which components and 'dose' of the components of the behavioral intervention are most effective. Factorial designs allow for an estimate of the individual effects of the components and whether the components interact with each other and are an efficient use of time, money or resources.

OBJECTIVE

The main aim is to identify which components or 'dose' of the components contribute to a reduction in systolic blood pressure over one year. We hypothesize that the most intense level will reduce systolic blood pressure from baseline to 12 months and that the effects will be additive.

Additionally, Reach Out will determine the impact of facilitated primary care provider appointment scheduling and transportation on primary care provider follow up of hypertensive patients treated in an urban, safety net ED. We hypothesize that subjects who receive primary care provider

appointment scheduling and transportation will have greater primary care provider follow up compared to subjects without primary care provider appointment scheduling and transportation.

IMPACT

Reach out leverages health system data to pragmatically identify subjects and test mobile health strategies to improve the access to and utilization of preventative health care services. By connecting ED patients to primary care providers, Reach Out may serve as a model for safety net health systems to improve chronic disease management in underserved communities.

Contributors

Lesli Skolarus
The University of Michigan

Casey Corches
The University of Michigan

Emily Champoux
The University of Michigan

William Meurer
The University of Michigan

A24

MATERNAL CHARACTERISTICS ASSOCIATED WITH CONSENT AND REASONS FOR REFUSAL IN A RANDOMIZED CONTROLLED TRIAL IN PREGNANCY

GAIL MALLETT

EUNICE KENNEDY SHRIVER NICHD MFMU NETWORK

Background/objective: Understanding the reasons for refusal can impact the design of recruitment strategies in future trials. We therefore sought to evaluate the characteristics associated with refusal to consent to a randomized trial in pregnancy.

Methods: This is a secondary analysis of an open-label multi-center RCT conducted by the Eunice Kennedy Shriver NICHD Maternal-Fetal Medicine Units (MFMU) Network in which women who were pregnant with one baby and had never given birth were randomly assigned to either induction of labor at 39 weeks of gestation or expectant management (i.e., forego elective delivery before 41 weeks of gestation). After initiation of the trial, the Data and Safety Monitoring Committee requested the addition of 3 fields to the screening log which already collected information about race/ethnicity: 1) maternal age, 2) type of insurance and 3) if refused to participate in the trial, the reason for refusal. Multivariable binomial and multinomial logistic regression, adjusted for each of the demographic variables on the screening log and hospital region, were used to estimate odds ratios (OR) and 95% confidence intervals (CI).

Results: The trial was conducted between March 2014 and August 2017. From August 2016 (start of additional data collection) to August 2017, 1,965 (28%) of the 7,089 eligible women consented to the trial. Consent rates were higher for non-Hispanic black women (41%, aOR 1.47, 95% CI 1.24-1.74) and lower for Asian women (16%, aOR 0.64, 95% CI 0.48-0.84), compared with non-Hispanic white women (24%). Women with government-assisted insurance or self pay had a higher consent rate (38%, aOR 1.55, 95% CI 1.34-1.79), compared with those with private insurance (22%). Consent rates decreased with increasing age. Among eligible women who refused participation in the trial and provided a reason, the reasons cited included wanting to be expectantly managed (85%), not wanting to participate in research (6%), objection of family and friends (5%), and wanting to be induced (3%). Not wanting to participate in research was more common in Asian women (aOR 2.41, 95% CI 1.44-4.04) and less common in Hispanic women (aOR 0.62, 95% CI 0.40-0.97), compared with white women. Asian women also were more likely to refuse because their family or friends objected (aOR 2.51, 95% CI 1.27-4.95) or because they wanted to be expectantly managed (aOR 1.74, 95% CI 1.30-2.32), compared with white women. Women with government-assisted insurance or self pay were more likely to refuse due to family or friends objecting (aOR 1.68, 95% CI 1.09-2.58) and were less likely to refuse because they wanted to be expectantly managed (aOR 0.60, 95% CI 0.51-0.71), compared with women with private insurance.

Conclusions: Consent rate and reasons for refusal were impacted by age, type of insurance, and race/ethnicity. These findings should be carefully considered as an aid for developing recruitment strategies that promote diverse representation of potential participant populations.

Contributors

Kim Hill
Eunice Kennedy Shriver NICHD MFMU Network

Eunice Kennedy Shriver NICHD MFMU Network

A25

DESIGNING A NON-INFERIORITY TRIAL: BAYESIAN VERSUS FREQUENTIST APPROACH

LYDIA FOSTER

MEDICAL UNIVERSITY OF SOUTH CAROLINA

Non-inferiority trials are designed to demonstrate that an experimental treatment is not worse than an established treatment by a specified margin. These trials are often designed using a frequentist approach; however, Bayesian approaches to non-inferiority trial design have recently been highlighted in the literature, including methods to select the non-inferiority margin, use of informative and non-informative priors, and decision rules for hypothesis testing. The NIDDK-funded Stent vs. Indomethacin for Preventing Post-ERCP Pancreatitis (SVI) Trial is being conducted to determine if indomethacin alone is non-inferior to the combination of indomethacin and prophylactic pancreatic stent placement (PSP) in patients undergoing high-risk ERCP who require PSP for the sole purpose of pancreatitis prevention. Using a frequentist paradigm, this trial was designed to have a maximum uninflated sample size of 1300 to assure 85% power to identify less than a 5% difference in post-ERCP pancreatitis (PEP) rates between the two treatment groups, assuming equal PEP rates of 9.7% in each arm. A Bayesian approach is used to redesign the existing trial and a simulation study conducted to investigate operating characteristics, including sample size, type I error, and power. This presentation will share the results of the simulation study, compare the two trial designs, and provide design recommendations for non-inferiority trials with a binary endpoint.

Contributors

Valerie Durkalski-Mauldin
Medical University of South Carolina

A26

IDENTIFYING SUBGROUPS OF CHILDREN WITH ASTHMA WHO ARE AT HIGHER RISK OF HAVING AN UNSCHEDULED MEDICAL CONTACT IN SEPTEMBER: A MIXED METHODS STUDY

REBECCA SIMPSON

UNIVERSITY OF SHEFFIELD

It has been proposed that the reported increase in the number of unscheduled medical contacts in September, among school-aged children with asthma, is caused by a viral challenge at the start of the school year. It has also been hypothesised that this challenge is compounded by children not taking their asthma medication over the summer holiday.

The aim of this research was to identify which subgroups of children with asthma were 'at risk' of having an unscheduled medical contact following the return to school in September.

A mixed methods approach was used to investigate this aim. The quantitative data came from the PLEASANT trial (Preventing and Lessening Exacerbations of Asthma in School-age children Associated with a New Term) a cluster-randomised intervention study. The PLEASANT data analysis informed the development of a two-stage qualitative study, before and after the school summer holidays. The qualitative results then informed further quantitative analyses of the PLEASANT data to validate the qualitative results. The various analyses were used to determine which subgroups of children with asthma were more 'at risk' of having unscheduled medical contacts in September.

Semi-structured qualitative interviews were undertaken with children with asthma aged 5-16 to explore their views as to why there are unscheduled contacts following the return to school. These results were used to identify initial subgroups of children who are 'at risk' of an unscheduled medical contact.

The qualitative results indicated that different levels of exercise could play a key role in the September increase in unscheduled medical contacts. Specifically, those children that have reduced levels of exercise over the summer holiday period could be at higher risk of having an unscheduled medical contact after their return to school, when their exercise levels increase through school sports. The qualitative results also indicated that those subgroups of children potentially most 'at risk', were those in the transition year from primary to secondary school, secondary school children and secondary school females.

The results from the qualitative study were tested through quantitative analysis of the PLEASANT dataset. Various methods were used to conduct the analysis including, logistic regression and negative binomial regression. This quantitative analysis results suggested that the subgroups of children with asthma who are potentially most 'at risk' of having an unscheduled medical contact in September were those with reduced exercise during the summer holidays, females, primary school children and those who were in the transition year from primary to secondary school.

In conclusion, it is suggested that a targeted intervention associated with an exercise regime over the summer holidays may be successful in reducing the number of school children experiencing unscheduled medical contacts in September.

Contributors

Steven Julious
University of Sheffield

Wendy Baird
University of Sheffield

A27

DEVELOPING A REMOTE STATISTICAL MONITORING PLAN

HEATHER MURRAY

UNIVERSITY OF GLASGOW

Background

Monitoring of clinical trials is necessary to ensure the protection of the study participants and the conduct of high-quality studies1. The aim of remote statistical monitoring is to carry out routine analyses of accumulating study data within a clinical trial to identify abnormal patterns, at individual study centres or groups of study centres, which might indicate deviations from the study protocol or from regulatory guidelines. The expectation is that remote statistical monitoring will reduce the need for and the cost of on-site monitoring.

Methods

A Remote Statistical Monitoring Plan (RSMP) should be developed at the beginning of the trial to specify the monitoring requirements. The RSMP should identify key data items to be monitored, the frequency of reports, describe the statistical methods to be used to identify abnormal patterns of data and the processes for escalation of potential issues identified. No single approach to monitoring is appropriate or necessary for every trial therefore the RSMP should be adapted to each trial based on the data integrity risks of the trial and the consequences these could have on the safety of the study population1. It is likely that the RSMP will have to be modified as the trial progresses.

Conclusions

Potential issues that remote statistical monitoring may flag include under reporting of serious adverse events or endpoints, missing data, errors in data (outliers, reporting of incorrect units for laboratory data, miscalibration of instruments used in the collection of data), non-compliance to study protocol, possible fraud or lack of understand of the protocol and delay in completing the Case Report Form (CRF) by on-site study staff. We will provide examples of output used in remote statistical monitoring reports and discuss the potential advantages and disadvantages of remote statistical monitoring compared to traditional on-site monitoring.

References

1U.S. Department of Health and Human Services, Food and Drug Administration. Oversight of Clinical Investigations – A Risk-Based Approach to Monitoring. Available at https://www.fda.gov/downloads/drugs/guidancecomplianceRegulatoryInformation/guidances/ucM269919.pdf

Contributors

Kirsty Wetherall

University of Glasgow, Institute of Health and Wellbeing, Robertson Centre for Biostatistics

Ian Ford
University of Glasgow, Institute of Health and Wellbeing, Robertson Centre for Biostatistics

A28

CHALLENGES IN CONDUCTING DART (S1609), A PIONEER BASKET TRIAL FOR RARE CANCERS IN A COOPERATIVE GROUP SETTING

MELISSA PLETS

FRED HUTCHINSON CANCER RESEARCH CENTER

Patients with rare cancers are vastly underrepresented in clinical trials, despite the fact that collectively rare tumors comprise 22% of all cancer diagnoses. Immune checkpoint blockade is a promising area of study in oncology, but clinical development has focused on more common tumor types. SWOG's DART (Dual Anti-CTLA-4 and Anti-PD-1 Blockade in Rare Tumors) trial aims to fill this void. Hundreds of rare solid cancers are eligible to enroll in one of 36 histologic cohorts, or the umbrella "not otherwise categorized" cohort. Each of the 36 histologic cohorts uses the same single-arm two-stage design for the endpoint response (RECIST confirmed complete or partial response by), with 6 patients in the first stage and 10 in the second (null=5%, alternative=30%). The trial activated on 1/13/17. As of 12/1/17, 148 patients have registered. This unique trial has provided a number of unique challenges. Three specific challenges are: triaging data for review, accurately predicting when cohorts should be closed to accrual, and the extremely high volume of email inquiries from sites.

Accrual has been 3x faster than expected, which has led to challenges in keeping up with data review. To facilitate review, we have ranked review priorities. Eligibility and response data are the top priority. Among response data, data review is further ranked to prioritize reviewing data most likely to impact the opening/closing of cohorts.

The first stage accrual is 6 patients. SWOG has no prior experience with nearly all of the histologic cohorts on the trial, and so it has been difficult to forecast cohort accrual. To further complicate the issue, it is SWOG standard practice to provide a 2-week notice prior to closing a cohort to accrual to allow any patients who were in the process of consenting the opportunity to register. As a result, we have over-accrued in all of the closed cohorts, some by as much as 30%.

Email volume is particularly high for this study for two reasons: 1) sites want to ensure that their patients' diagnoses are confirmed to be rare and eligible for the trial and 2) accrual to cohorts varies widely and cohorts are opening and closing frequently. We have addressed the first issue by sheer manpower with the study team spending several hours a day responding to these questions. We have addressed the second issue by developing a public website with accrual and closure status updated daily.

Contributors

Edward Mayerson
Fred Hutchinson Cancer Research Center

Megan Othus
Fred Hutchinson Cancer Research Center

A29

DETERMINING WEAR TIME VALIDATION ALGORITHMS AND CUTPOINTS: CHALLENGES TO USING ACCELEROMETRY IN A BEHAVIORAL RANDOMIZED CLINICAL TRIAL OF ADOLESCENTS WITH CONGENITAL HEART DISEASE

JAMIE JACKSON

NATIONWIDE CHILDREN'S HOSPITAL

INTRODUCTION: Accelerometry is the gold standard for measuring physical activity. Investigators decide the wear time validation algorithm, which makes determinations about intervals of inactivity vs. disuse, as well as cutpoints for moderate and vigorous physical activity (MVPA) to obtain the most valid data. These decisions depend on multiple factors, including the age and physical activity patterns of the population. Challenges in making these decisions are highlighted when implementing an NIH-funded behavioral RCT to increase MVPA among 15-18 year-olds with critical congenital heart disease (CHD). Preliminary baseline data on accelerometry in this population demonstrates how wear time algorithm and cutpoint choice impacts interpretation of accelerometry data, which has implications for participant eligibility and validity of study outcomes.

METHODS: Currently, 70 adolescents (15-18 year-olds) with critical CHD are being recruited for baseline assessment to determine eligibility for randomization into a physical activity lifestyle intervention. To be eligible, in part, participants must engage in <60 min on average of MVPA per day as measured by an Actigraph accelerometer (wGT3X-BT) worn around the waist. Wear time must be >4 days (10 hrs/day; 1 weekend day), to determine participant eligibility. Data is analyzed using Actilife 6.13.3. Once randomized, participants are asked to wear the accelerometer midpoint (9 weeks) and at study conclusion (21 weeks) regardless of active or comparison arm assignment. Outcomes include the average number of minutes spent in MVPA (primary) and sedentary activity (secondary). Two wear time validation algorithms are available on Actilife 6, including Troiano (2007) and Choi (2011), as well as several cutpoints, which have been validated for healthy children and adults, including Freedson (2005; children and adults) and Evenson (2008; children). For studies with adolescent CHD survivors, wear time algorithms are not typically reported and both Freedson and Evenson cutpoints have been used.

OBSERVATIONS: Baseline accelerometer data has been collected on 5 participants with an average of 6.4 days of wear (range 5-7 days). When using Troiano to define wear time, 2 participants had 1 day that was considered a “non-wear” day in contrast to when using Choi. This resulted in a discrepancy of 593.8 minutes of unaccounted sedentary behavior across the 5 participants. When using Choi for wear time, the following average number of minutes spent in MVPA were derived across the 5 participants: 183.1 (Freedson child), 28.9 (Evenson child), and 39.4 (Freedson adult). Abstract presentation will include 20+ participants.

CONCLUSIONS: When working with a population in which wear time algorithms and cutpoints have not been validated, investigators must make educated guesses. The current preliminary data suggest that these choices may significantly impact data validity, including which participants are deemed eligible based on both the number of days the device is worn and average time spent in

MVPA. For CHD, significant variability in physical activity has been reported across studies, which may be largely due to the use of different wear time algorithms and cutpoints for MVPA. This may lead to erroneous conclusions about physical activity in CHD survivors that could have downstream consequences for patient care.

Contributors

Carine Leslie
Nationwide Children's Hospital

Kathryn Vannatta
Nationwide Children's Hospital

Jennifer Cotto
Nationwide Children's Hospital

Joseph Rausch
Nationwide Children's Hospital

A30

MY DATA, YOUR DATA, WHERE'S THE DATA?

STACEY SLONE

UNIVERSITY OF KENTUCKY

Reproducibility of results is a hot topic in clinical trials. While the statistical methods and results are generally the focus of such discussions, having the exact data used for the analyses is an integral aspect of reproducibility. During internal discussions concerning reproducibility, the various approaches being implemented in regards to data download, storage and final data creation were concerning. Each statistician had an individual strategy but very little commonality. The inefficient processes being utilized at our institution were increasing the time burden on faculty and staff. In order to facilitate the efficiency of trial maintenance and the reproducibility of results, we developed a strategy to minimize the repetitive steps and harmonize the data creation and storage within the unit.

The plan centers on data collected in OnCore, a commonly used clinical database management system (CDMS) used at many cancer centers in the United States. While OnCore allows for the download of current clinical trial data into SAS, the system is a multi-step process. Since cancer centers can customize OnCore eCRFs to specify explicit data points to be collected for a particular clinical trial, the data download system has to be very general. This generalizability can cause issues within the SAS code provided by OnCore when downloading data from multiple eCRFs. Our strategy begins by highlighting download selections in OnCore, creates a harmonious directory structure and finally incorporates SAS macros and %INCLUDE statements to establish the process. Applied wisely, the current process can circumvent many chaotic situations in data handling, reduce the time necessary to accomplish certain tasks by automation and standardization, eliminate duplicated effort and guarantee a level of quality in our current data reporting.

Although our specific strategy focuses on data downloaded from OnCore, the approach and practices can be generalized to other CDMS. Lessons learned and future directions will be presented. As noted business historian, Alfred D Chandler, Jr, stated, "Unless structure follows strategy, inefficiency results."

Contributors

Emily Dressler
Wake Forest School of Medicine

Rani Jayswal
University of Kentucky

Meng Liu
AbbVie Inc

A31

ANALYSIS OF STUDIES WHEN RESCUE THERAPY CENSORS DATA

HEIDI CHRIST-SCHMIDT

STATISTICS COLLABORATIVE, INC.

In a typical well-designed, carefully executed, randomized clinical trial, all participants are followed until the end of the study and the primary outcome is assessed for every participant. Some metric summarizing the difference between the treatment groups is calculated to evaluate the efficacy of each test intervention. An important current topic in biostatistics deals with methods for handling missing data. In the language of Little and Rubin, missing values are categorized as missing completing at random, missing at random, or missing not at random. Many rigorous statistical methods are available for data that are missing completely at random or missing at random. More problematic are those whose data are missing not at random. Often participants in trials whose primary outcome is missing comprise a select, rather than a random, group. Those missing primary outcome data are often, on average, either more ill or more healthy than the average participant. In some situations, the data from such participants can be modeled as missing at random.

In certain situations, trial protocols describe "rescuing" participants whose clinical status deteriorates. Such rescue, while clinically necessary, may distort the assessment of the effect of the experimental drug because rescue produces data that are missing in a way that precludes estimation of what would have happened had the participants remained on study drug. Common examples of such situations are studies of interventions for chronic pain, chronic arthritis, and schizophrenia. In these cases, rescue occurs in one direction (pain too high; symptoms of schizophrenia or arthritis too severe). Studies whose outcomes are laboratory measurements, such as serum potassium or fasting blood glucose, may require rescue if the results are either too high or too low.

This presentation describes a method for assessing the effect of drug in a way that is statistically rigorous but that recognizes the clinical imperative of maintaining safety. The method relies on carrying the last rank (rather than last value) forward.

Contributors

Yingxin Hou
Defense Health Headquarters

Janet Wittes
Statistics Collaborative, Inc.

A32

EXPLORING THE IMPACT OF THE FALSE DISCOVERY RATE TO CORRECT FOR MULTIPLE COMPARISONS

ZHAOMIAN LI

JAEB CENTER FOR HEALTH RESEARCH

BACKGROUND: It is not uncommon for a randomized clinical trial protocol to produce big datasets that are used for a large number of secondary analyses. Investigators are often resistant to formally correcting for multiple comparisons in secondary analyses because of the reduction in power, especially when controlling for the family-wise error rate (FWER). A well-established alternative to FWER is to control the false discovery rate (FDR).

METHODS: Secondary analyses from 6 different diabetes-related randomized clinical trials were retrospectively corrected for multiple comparisons using 3 prominent FDR methods: Benjamini-Hochberg FDR (FDR), adaptive FDR (aFDR), and positive FDR (pFDR). The analysis included the exact same list of statistical comparisons that were performed in the original data packet for each study. The corrected p-values were compared with the original uncorrected p-values to determine the impact on statistical power.

RESULTS: A total of 432 p-values from secondary analyses across the 6 studies were analyzed, with the number of comparisons per study ranging from 41 to 106. Among the 112 comparisons where the nominal p-value was <0.01, there was not a single case where the aFDR-corrected value was >0.05 and only one case where the FDR-corrected value was >0.05. Among the 70 cases where the nominal p-value was between 0.01 and 0.05, the corrected value was still <0.05 49% and 71% of the time for FDR and aFDR, respectively.

Results varied substantially by study according to the distribution of nominal p-values. The proportion of nominal p-values below 0.05 ranged from 2% to 64%. This affected statistical power, with the percentage of nominal p-values between 0.01 and 0.05 that were still <0.05 after corrections ranging from 8-100% for FDR and 42-100% for aFDR. pFDR was problematic, often failing to produce an adjustment when it cannot estimate a positive number of true null hypotheses.

CONCLUSIONS: For studies with a large number of secondary or exploratory analyses, FDR is a reasonable compromise between controlling FWER and ignoring the problem of multiple comparisons altogether. The effect on statistical power depends on the proportion of true null hypotheses, which can vary substantially by study. The results from 6 RCTs in diabetes suggest that FDR is usually more powerful than a common informal procedure of shifting the threshold for statistical significance to 0.01. aFDR is less conservative than FDR and pFDR can be problematic when the method cannot estimate a positive number of true null hypotheses.

Contributors

A33

INTEGRATED DRUG TRACKING MODULE IN EDC-CTMS FOR OPERATION QUALITY ASSURANCE IN A LARGE MULTICENTER TRIAL

WENLE ZHAO

MEDICAL UNIVERSITY OF SOUTH CAROLINA

AtRial Cardiopathy and Antithrombotic Drugs In Prevention After Cryptogenic Stroke (ARCADIA) is a multicenter, randomized controlled trial to test the hypothesis that apixaban is superior to aspirin for the prevention of recurrent stroke in patients with cryptogenic ischemic stroke and atrial cardiopathy. Special challenges for study drug tracking in the ARCADIA trial include: 1) large multicenter trial with 120 sites; 2) slow recruitment with an estimated 4 subjects per site per year; 3) long term follow-up (1.5 - 4 years) with study drug resupply every 3 months of which a mixture of in-person and telephone/mailing visits will be used; 4) adjustable apixaban dose (5 mg vs 2.5 mg) based on bio-characteristics assessed every 3 months; and 5) double-dummy treatment masking with study drug kits containing active apixaban and matching aspirin placebo or active aspirin and matching apixaban placebo.

To ensure that the correct study drug kit is available at the time of randomization and resupply while mitigating study drug waste, a comprehensive study drug tracking module has been developed and integrated into a web-based EDC-CTMS system. This module manages the drug bottle packing, drug kit assembly, drug shipping from the central pharmacy, drug receipt by the site, drug kit assignment to a subject, drug kit dispensing, and drug removal from site inventory due to expiration/damage. The study drug module is integrated with the site management module, subject randomization module, subject study progress module, and Case Report Form data management module, so that study drug requests are automatically posted for the central pharmacy. This design ensures that each site currently released to enroll has the minimal study drug inventory to avoid loss of subjects due to drug unavailability at the time of randomization. It also ensures the correct study drug kit with the correct treatment arm and dose is available for each subject when a resupply drug kit is needed. Barcode scanning technology is used for drug bottle labels, drug kit labels, and bulk drug barrels to augment central pharmacy quality assurance procedures. Likewise, verification codes on the individual drug kits are utilized to ensure the proper drug kit is dispensed to the subject.

Contributors

Catherine Dillon
rileycp@musc.edu

Kieth Pauls
paulsk@musc.edu

Caitlyn Meinzer
ellerbcn@musc.edu

A34

EMPIRICAL ASSESSMENT OF TRIAL FUTILITY IN MULTICENTER CLINICAL TRIALS

ZHIBAO MI

VA COOPERATIVE STUDIES PROGRAM COORDINATING CENTER

Stopping a futile clinical trial early can save limited resources and prevent unnecessary exposure to adverse effects. However, if a trial is stopped inappropriately because of a premature analysis, its conclusion may be informed by inadequate and unbalanced information. Therefore, choosing the right time and frequency for interim analyses is challenging and very important. Current methods to assess futility are mostly based on stochastic curtailment, predictive power, and group sequential design, but these methods all assume that the patient data collected before the interim analysis is homogenous and similar to those after the analysis. In practice, this assumption may not always be true. Investigators may not be as proficient at executing study protocol earlier in the study and inadvertently provide different care than later. Furthermore, the study protocol could be amended before and after the interim analysis. Certain diseases' prognoses vary with time. For multicenter trials, some centers start enrollment earlier than others due to logistic reasons, and consequently, enrollment may not be balanced among all participating centers at the interim analysis, which would be biased by any center-related effects. In this report, we retrospectively examine multicenter clinical trials with different types of endpoints (binary, continuous, and time-to-event), including completed trials with and without treatment effects, as well as trials stopped early due to statistical, safety, and operational reasons. We then perform interim analysis for futility at different time points using conditional probability, predictive probability, and futile boundary. The futility assessment is focused on time and center effects to evaluate sample size and information accumulated to maintain stable time variation. In addition, we use bootstrapping resampling procedure to provide more robustness when interpreting single estimate from a trial analysis. The results of this assessment may help clinical trialists better plan for when to perform a futility analysis.

Contributors

Kelsey Alexovitz
VA Cooperative Studies Program Coordinating Center

Xiaoli Lu
VA Cooperative Studies Program Coordinating Center

Kousick Biswas
VA Cooperative Studies Program Coordinating Center

Joseph Collins
VA Cooperative Studies Program Coordinating Center

A35

REDUCING UNNECESSARY ANTIBIOTIC PRESCRIBING IN PRIMARY CARE: A POINT-OF-CARE ELECTRONIC INTERVENTION USING ELECTRONIC HEALTHCARE RECORDS (REDUCE TRIAL)

JAMIE SOAMES

CLINICAL PRACTICE RESEARCH DATALINK (CENTRE OF THE MHRA)

Introduction: Unnecessary use of antibiotics in primary care is contributing to the development of antimicrobial resistance. We evaluated the effectiveness of multi-component interventions delivered into routine care through electronic healthcare record (EHR) systems at reducing antibiotic prescribing for self-limiting respiratory tract infections (RTI).

Methods: Seventy-nine general practices contributing to the UK Clinical Practice Research Datalink (CPRD) were cluster-randomised to intervention or control trial arms. Multi-component interventions were introduced using a webinar and comprised 1) antibiotic prescribing Decision Support Tools triggered in the EHR system during RTI consultations including patient information sheets used to address patients' concerns about their illness and the self-management of symptoms 2) monthly primary care clinic-specific feedback data for respiratory consultations and antibiotic prescribing in comparison with the preceding 12 months. All outcomes were obtained from primary care EHRs in the CPRD GOLD database. The primary outcome was the rate of antibiotic prescriptions for self-limiting RTI over the 12-month intervention period. Analysis was by Poisson regression of patient-level data, including general practice as a random effect, adjusting for age-group, gender, comorbidity, region, study quarter, period of randomisation and baseline antibiotic prescribing. A process evaluation was conducted using an online questionnaire and data on utilisation of Decision Support Tools.

Results: There were 41 intervention trial arm practices (323,155 patient-years) and 38 control trial arm practices (259,520 patient-years). At baseline, the two groups were similar in terms of age, gender, co-morbidity and antibiotic prescribing rates. There were 98.7 AB prescriptions for RTI per 1,000 patient-years in the intervention trial arm (31,907 AB prescriptions) and 107.6 per 1,000 in the control arm (27,923 AB prescriptions); adjusted AB prescribing rate ratio (RR) 0.88 (95% confidence interval 0.78 to 0.99, P=0.040). There was no evidence of effect in children less than 15 years (RR 0.96, 0.82 to 1.12, P=0.632) or adults aged 85 years and older (RR 0.97, 0.79 to 1.18, P=0.742), AB prescribing was reduced in adults 15-84 years (RR 0.84, 0.75 to 0.95, P=0.006). One antibiotic prescription was avoided for every 62 (95% confidence interval 40 to 200) patients aged 15-84 years per year. Analysis of trial data for 11 safety outcomes, including pneumonia and peritonsillar abscess, showed no evidence that these might be increased as a result of intervention.

Conclusion: Antibiotic prescribing was reduced by 16% in adults aged 15 to 84 but we did not find evidence of a reduction in children or those aged 85 years or over. Future interventions should consider the requirements of different age-groups. In addition, providing prescribing data to primary care physicians as a single intervention may help to raise awareness of antibiotic prescribing habits in the clinic, potentially resulting in better management of antibiotic prescriptions and providing a light intervention which might be scaled up to a wider population.

Finally, this trial provides further evidence that EHR systems can be used successfully both to deliver an intervention and to evaluate outcomes.

Contributors

Kirin Sultana
Clinical Practice Research Datalink (centre of the MHRA)

Mark Wright
Clinical Practice Research Datalink (centre of the MHRA)

Judith Charlton
King's College London, School of Population Health and Environmental Sciences

Toby Prevost
Imperial College London, Faculty of Medicine, School of Public Health

Mark Ashworth
King's College London, School of Population Health and Environmental Sciences

A36

SITE INITIATION TRAINING – WHAT CAN BE LEARNED FROM EVALUATING THIS PROCESS FROM MULTICENTRE TRIALS?

KAREN INNES

UNIVERSITY OF ABERDEEN

Background: Clinical trials are expensive to design and run. Site training is a fundamental part of opening a site to recruitment and monies for completing these tasks are included within funding applications. Different modes of training are available including face-to-face on-site training, teleconference, video conference or investigator days: each mode has implications for the budget, trial office staff and site research staff.

Currently within our Clinical Trials Unit (CTU) we use different modes of training, the most common being face-to-face on-site training (particularly during site set-up). We have not previously routinely evaluated site initiation training and have therefore started eliciting views and opinions of research staff at sites with regards to the mode of delivery and the training materials provided within large, multicentre randomised controlled trials in the UK.

Method: A purpose designed evaluation form is used to evaluate the mode of training and training materials used. Evaluation forms are given to the study team on the day of the initiation training, with the option to complete immediately or return by email at a later date. Data is being collected from five UK-funded National Institute for Health Research Health Technology Assessment Programme trials within our CTU.

Early data collection has commenced in the C-GALL Trial (A randomised controlled trial comparing laparoscopic cholecystectomy with observation/conservative management for preventing recurrent symptoms and complications in adults with uncomplicated symptomatic gallstones). As part of site initiation, training is provided by the C-Gall Trial Manager to site research staff during on-site face-to-face visits. Each visit takes around three hours and the Principal Investigator, research nurses and any other site staff are encouraged to attend. Each training session follows the same agenda, and uses trial-specific training materials which include sections on identifying eligible patients, audio recording, recruitment, follow-up, safety and breaches.

Results: Of the evaluation forms we have to date, 100% of research staff chose on-site face-to-face training as the preferred mode of training. Comments about this included: “Opportunity to engage in constructive discussion re the trial”, “Facilitates better communication”, “I am a visual person, I like to see paper copies of materials and practical demonstrations”. Research staff were also asked about the helpfulness of the training materials provided and over 87% rated these as excellent or very good. Comments about the materials provided for this study and training session in comparison to other studies included: “Very specific to each aspect/role. Very clear and user friendly”, “On par with other studies”, “Very favourably. Excellent, professional looking materials, well presented. Reassuring”.

Implications for practice: The mode of training delivery has implications for both CTU staff time and budget. Further evaluation of different modes of site training including materials required would better inform the preparation of the training budget and provide evidence on the most efficient mode of training site staff for clinical trials.

Data collection is on-going across a number of trials which utilise different methods of site initiation. We will present updated results and discuss lessons learned.

Contributors

Kath Starr
University of Aberdeen

A37

DEVELOPMENT OF A STANDARDISED SET OF METRICS FOR MONITORING SITE PERFORMANCE IN MULTICENTRE RANDOMISED TRIALS: A MIXED METHODS STUDY

DIANE WHITHAM

UNIVERSITY OF NOTTINGHAM

Background

Site performance is key to the successful delivery of large multicentre randomised trials. Measures of site performance should deliver meaningful, actionable information that can be used to monitor sites and initiate remedial action, if necessary. A standardised set of clear and accessible summaries of site performance could facilitate the timely identification and resolution of potential problems, minimising their impact. The aim of this study was to identify and agree a core set of key performance metrics and create a simple reporting tool for managing multicentre trials.

Methods

We used a comprehensive, mixed methods approach to identify potential metrics and to achieve consensus about the final set, adapting methods that are recommended by the COMET Initiative for developing core outcome sets in health care (www.comet-initiative.org/).

Firstly, we conducted a systematic search for studies describing ways of measuring individual site performance in multicentre randomised trials using the Cochrane Library, five biomedical bibliographic databases and Google Scholar. Data on study quality and content were extracted independently by two reviewers. We also held three focus group discussions of UK-based stakeholders (10-11 per group) to identify site performance metrics that are or could be measured routinely in randomised trials. Data were recorded, transcribed and analysed thematically using NVivo 11 qualitative data analysis software.

Performance metrics identified from the systematic search and focus groups were used to create an online Delphi survey. We invited respondents to score each metric for inclusion in the final core set over three survey rounds. Metrics scored as “critical” by at least 70% and “unimportant” by <15% of survey respondents were retained for discussion at a consensus workshop of representatives from key UK-based stakeholders.

Results

We identified 119 performance metrics from 25 eligible studies in the systematic literature search, and 19 from the 32 participants in the three focus groups. Metrics were categorised as relating to site potential, recruitment, retention, data collection and quality, trial conduct and trial safety.

Round 1 of the Delphi survey presented 28 performance metrics, and a further six were added in round 2, following participant feedback in round 1. Of 294 UK-based stakeholders who registered for the Delphi survey, 211 completed all three rounds.

Fifteen metrics were retained following round 3. These were discussed and voted on at the consensus workshop. Consensus was reached on a final set of eight key performance metrics. These were in three domains: (1) recruitment and retention, (2) data quality, and (3) protocol compliance. We have created a simple tool for effective visual reporting of the metrics which is freely available at http://www.nottingham.ac.uk/nctu/other-research/other-research.aspx and could be used alongside existing systems.

Conclusions

By using robust methods to achieve consensus, we have established a core set of metrics for measuring performance of sites in multicentre randomised trials. These metrics could improve efficient trial conduct by enabling researchers to identify and address problems before trials are adversely affected. Future work could evaluate the effectiveness of using the metrics and reporting tool.

Contributors

Julie Turzanski
NCTU, University of Nottingham

Alan Montgomery
NCTU, University of Nottingham

Kate Walker
NCTU, University of Nottingham

Lucy Bradshaw
NCTU, University of Nottingham

Shaun Treweek
HSRU, University of Aberdeen

A38

DUAL-AGENT DOSE ESCALATION METHODS FOR PEDIATRIC ONCOLOGY CLINICAL TRIALS

CLEMENT MA

DANA-FARBER/BOSTON CHILDREN'S CANCER AND BLOOD DISORDERS CENTER

Phase 1 clinical trials aim to identify the optimal dose for the therapeutic agent that balances patient safety and potential efficacy. Cancer therapies that include two or more agents may increase efficacy as well as toxicity. Many adaptive dose-escalation designs have been proposed for trials of combination therapies. These designs can better assign dose combinations near the maximum tolerated dose combination (MTDC) to enrolled patients but require significant resources to design and monitor. Hence, relatively few adult oncology trials have used these designs, and to our knowledge, none have been used in pediatric trials. To motivate the use of adaptive designs in pediatric oncology, we performed a simulation study to compare the performance of dual-agent dose-escalation methods in a pediatric oncology framework.

We selected four Bayesian methods including the: (1) Partial Order Continual Reassessment Method; (2) Bayesian Optimal Interval design; (3) Conaway, Dunbar, and Peddada method; and (4) Copula method for our study. For comparison, we also included the commonly used 3+3 rule-based design assuming a pre-specified ordering of dose combinations. We designed 7 simulation scenarios with a restricted number of dose combinations and low total sample size (N=24) to reflect the realities of pediatric trials. We performed 2,000 simulated trials per scenario for each method and compared the methods across six metrics. Overall, all adaptive methods had a similar performance across all metrics. The average recommendation rates for the true dose combination ranged from 37% to 43%. The average proportion of patients receiving a dose greater than the MTDC ranged from 29% to 33%, which is near our target toxicity level of 30%. As expected, the conservative 3+3 design had lower recommendation rates than the adaptive designs. Adaptive designs represent a safe and effective way for dual-agent dose escalation trials in children.

Contributors

Derek Shyr
Dana-Farber/Boston Children's Cancer and Blood Disorders Center

Hasan Al-Sayegh
Dana-Farber/Boston Children's Cancer and Blood Disorders Center

Dongjing Guo
Dana-Farber/Boston Children's Cancer and Blood Disorders Center

A39

ANALYZING TREATMENT EFFECTS AND MODERATORS IN RANDOMIZED PRE-POST CLINICAL TRIALS

JOSEPH RAUSCH

NATIONWIDE CHILDREN'S HOSPITAL, OHIO STATE UNIVERSITY MEDICAL CENTER

Primary analyses in randomized clinical trials (RCTs) are often based on pre-post data where baseline variables are measured and potentially of interest as covariates or moderators (see, e.g., Kraemer et al, 2002, Archives of General Psychiatry), and the pretest is a special case of a baseline variable that is collected on the outcome measure. Analysis of covariance with prespecified covariates, often including the pretest, is a common and useful approach for analyzing treatment effects on continuous outcomes based on pre-post data from RCTs. Furthermore, extensions of analysis of covariance that allow for non-parallel regression lines between the moderators and the outcome across the treatment groups (e.g., Rogosa, 1980, Psychological Bulletin) are often used to analyze moderators within this same context.

An important problem with these methods is that, although these approaches are advantageous with complete pre-post data from an RCT, they are typically sub-optimal when some of the outcome data are missing. Specifically, these methods are generally used in the context of the general linear model, and thus require researchers to drop RCT participants with missing outcome data from the analysis. Whereas such an approach can be appropriate when data are missing completely at random, it is more common to expect that the actual missing data mechanism when analyzing treatment effects or moderators from pre-post data within an RCT is missing at random or not missing at random (see, e.g., Schafer & Graham, 2002, Psychological Methods, for more information on missing data mechanisms).

The primary goal of the current study is to compare and contrast potential approaches that allow for more flexible missing data mechanisms when analyzing treatment effects and moderators from pre-post data in RCTs. In general, maximum likelihood and multiple imputation methods are considered the state-of-the-art for analysis when data are missing (Schafer, 2002, Psychological Methods). Nevertheless, little research appears to be available on these methods when analyzing treatment effects and moderators from pre-post data in RCTs. Consequently, the current study investigates (a) an extension of the mixed model approach for pre-post data (see, e.g., Winkens et al, 2007, Contemporary Clinical Trials) for maximum likelihood analysis and (b) a multiple imputation approach to analysis of covariance and its non-parallel regression line extension for the analysis of treatment effects and moderators.

These models are introduced and compared to determine their underlying operating characteristics within the practice of the analysis of RCTs from pre-post data, focusing on aspects such as (a) model specification for the analysis of treatment effects, (b) dealing with interactions terms within these models, (c) the effect of random covariates on their statistical properties, (d) the small-sample properties of these methods, and (e) the potential for employing auxiliary variables to account of missing data within these approaches. Ultimately, it is shown that these methods have their own

relative advantages and disadvantages for the analysis of treatment effects and moderators from pre-post data within an RCT.

Contributors

A40

DESIGN AND OPTIMAL USE OF THE EXPERTISE-BASED RANDOMISED CONTROLLED TRIAL DESIGN

JONATHAN COOK

UNIVERSITY OF OXFORD

Randomised controlled trials (RCTs) are widely recognised to be the most rigorous way to test new and emerging clinical interventions. When the interventions under evaluation are two different surgical procedures, however, surgeons are required to be proficient in the different surgical approaches to take part in such a trial. It is often the case that even where surgeons can perform both trial surgical procedures, they have a preference and/or have more expertise in one of the procedures. The expertise-based trial design, where patients are randomised but participating surgeons only provide the procedure in which they have greatest expertise, has been proposed to overcome this problem. When an expertise-based design should be used remains unclear, however, and such approaches may be more suited to addressing specific questions.

In this presentation we will introduce the expertise based design and its use in practice, and present the findings of a depth qualitative investigation of surgeons and methodologists perspectives on the optimal conditions for its design and use. Insights gained suggest that the expertise-based trial design has significant potential to increase surgeon participation in trials; however, there are a number of ethical and logistical issues which have precluded its wider use to date (including methodological issues such as the concerns about potential for bias in the different arms of the trial and logistical issues such as the need to "transfer" patients between surgeons). Particularly suitable conditions include those where the procedures under evaluation are substantially different, where they are routinely delivered by different health professional/surgeons with clear proficiencies in each; and contexts in which a multiple surgeon model is operating and the trust between the patient and surgeons can be suitably protected.

Contributors

Marion Campbell
University of Aberdeen

Katie Gillies
University of Aberdeen

Zoe Skea
University of Aberdeen

A41

DESIGNING TWO- OR THREE-ARM RANDOMIZED PHASE 2 SELECTION TRIALS USING A MARGIN OF PRACTICAL EQUIVALENCE

HAKIM-MOULAY DEHBI

CRUK & UCL CANCER TRIALS CENTRE

Background:

In randomized phase 2 selection trials, multiple treatments are compared in order to select the most appropriate one(s) for a phase 3 trial against the standard of care or a placebo. In practice such trials are most often designed and analysed, using a "pick-the-winner" strategy where only efficacy is considered to compare treatments. The treatment that is superior in terms of efficacy by any margin is selected for further testing. Design-wise, popular "pick-the-winner" approaches include the one-stage Simon et al.'s randomised design [1] and the two-stage Thall et al.'s randomized designs [2-3].

Considerations other than efficacy, such as ease of administration, cost, toxicity and quality of life should be part of the decision-making process in a phase 2 selection trial. This approach better reflects real-life situations. To make this feasible, Sargent and Goldberg [4] introduced the concept of a margin of ambiguity to retain treatments that are slightly less efficacious than the winning treatment as potential candidates for a phase 3 trial. The authors considered the case of 2-arm trials and a few combinations of design parameters in their paper.

Objective / Method:

In this work, we extended the concept of the margin of ambiguity (renamed the margin of practical equivalence) to design randomized selection trials for an exhaustive set of parameters based on exact binomial probabilities. We considered the following combinations of parameters:

1/ two- or three-arms trials;

2/ the probability of success in the superior arm varying between 0.15 and 0.95 by step of 0.05;

3/ the difference in success probability between the superior arm and the other arm(s) of either 10 or 15 percentage points;

4/ the margin of practical equivalence of either 2.5 or 5 percentage points.

Results / Output:

We derived the probability of selecting the superior treatment (referred to as Lambda) for the whole set of combinations of design parameters described above. Based on these probabilities, we developed a user-friendly online tool to design randomized phase 2 selection trials with a margin of practical equivalence. After defining their four design inputs, the users can use our tool to obtain:

1/ Lambda given a fixed sample size; or

2/ the sample size that corresponds to a certain Lambda.

The tool can be found at (see figure attached):

https://hakdehbi.shinyapps.io/randomised_phase_2_margin_equiv/

[1] Simon, R., R.E. Wittes, and S.S. Ellenberg, Randomized phase II clinical trials. Cancer Treat Rep, 1985. 69(12): p. 1375-81

[2] Thall, P.F., R. Simon, and S.S. Ellenberg, Two-stage selection and testing designs for comparative clinical trials. Biometrika, 1988. 75(2): p. 303-310

[3] Thall, P.F., R. Simon, and S.S. Ellenberg, A two-stage design for choosing among several experimental treatments and a control in clinical trials. Biometrics, 1989: p. 537-547

[4] Sargent, D.J. and R.M. Goldberg, A flexible design for multiple armed screening trials. Statistics in medicine, 2001. 20(7): p. 1051-1060

Word count: 467

Contributors

A42

A COMPREHENSIVE CUSTOMIZABLE SITE READINESS CHECKLIST FOR EFFICIENT SITE ACTIVATION IN A CLINICAL TRIAL

JULIA COLLINS

THE EMMES CORPORATION

Successfully launching a clinical trial is a complex endeavor that requires the timely coalescence of several moving parts. Clinical trials often launch later than expected due to the challenge of starting a study on time, particularly when multiple sites are involved. This delay in study start can have many implications, from consuming limited budgets and breaking critical timelines, to a potentially dramatic effect on patient health outcomes and public policy. Given these implications, it is critically important for trial sponsors to utilize all available tools to assist them in launching a study on time. One of the most crucial components of study start-up is activating each trial site to begin enrolling participants in a timely manner. Unfortunately, sites often lag behind in completing the numerous tasks required prior to site activation, whether due to competing priorities, disorganization, unclear expectations, staffing issues, or a combination of these and other factors.

For over ten years, the National Institute on Drug Abuse (NIDA) Clinical Coordinating Center (CCC) has supported the National Drug Abuse Treatment Clinical Trials Network (CTN), which tests substance use treatment interventions in multi-site clinical trials. During this time, the CCC has developed tools and thorough procedures to maximize the efficiency and timeliness of study start-up, including the use of a comprehensive and customizable Site Readiness Checklist to prepare study sites for activation. This Checklist contains a list of all tasks the site must complete before becoming eligible for site activation and includes various functional areas, such as required regulatory tasks and documentation, site staffing and training, receipt and storage of study supplies and medication, and preparation of local Standard Operating Procedures (SOPs). During the early start-up phase, the CCC modifies the Site Readiness Checklist template to align with the specific needs of each individual study. When the timing is appropriate, the CCC distributes the Checklist to study sites and provides sites with a detailed explanation regarding content and use of the checklist, with all expectations for completing the Checklist. It is important that the Checklist is provided to sites early enough to promote the timely completion of start-up tasks (i.e., prior to National Training), but not so early that receipt of the Checklist overwhelms an understaffed or unprepared site. The CCC receives and reviews each submitted Checklist weekly to monitor the site's progress, then provides feedback to the site staff regarding pending tasks and any potential issues to focus them on key activities. The CCC also utilizes the Checklist to efficiently schedule Site Initiation Visits (SIVs) as close as possible to the time when the site is fully prepared for activation. Once the SIV and all pending action items (including those on the Checklist) are completed, the site becomes eligible for activation to begin enrolling participants. In the CTN, use of the Site Readiness Checklist has maximized the efficiency of the site activation process by setting clear expectations and timelines for sites, thereby facilitating further success in meeting budget and timeline expectations throughout the duration of the trial.

Contributors

Dikla Shmueli-Blumberg
The Emmes Corporation

Eve Jelstrom
The Emmes Corporation

Matthew Wright
The Emmes Corporation

Robert Lindblad
The Emmes Corporation

A43

THE CENTRAL IRB EXPERIENCE IN A MINIMAL RISK OBSERVATIONAL STUDY

LETITIA PERDUE

WAKE FOREST UNIVERSITY HEALTH SCIENCES

The SPRINT Alzheimer's, Senior and Kidney (SPRINT ASK) continues the original SPRINT protocol and will obtain critical follow-up data for the cognitive and kidney components of SPRINT. This minimal risk study includes one additional in-person clinic visit on approximately 7000 participants.

The SPRINT ASK grant was funded for only one year; thus, start-up at the nearly 100 SPRINT clinics had to be expeditious. Due to the tight timeline and the nature of this minimal risk study, SPRINT ASK was an ideal pilot study for the upcoming requirement by NIH to use a single Institutional Review Board (IRB) for multi-site research.

The SPRINT ASK leadership encouraged the use of the Coordinating Center as the central IRB for all non-VA clinics. Some clinics chose not to use the central IRB for a variety of reasons; however, the SPRINT ASK Coordinating Center serves as the central IRB for nearly half of the clinics throughout the United States and Puerto Rico. This presentation will review the initial start-up steps and things to consider prior to establishing a central IRB.

Reliance on Wake Forest to serve as the Central IRB was completed either through the Smart IRB system or by signing an Authorization Agreement at the two institutions. While several of the institutions were part of the SMART IRB agreement, few had used the system to indicate willingness to rely on another institution.

The study team worked with the Wake Forest IRB to develop a site submission packet. Staff at the clinics completed a site packet that included: Site Submission Form to address basic questions about participant population and research practices and delegations at the clinic; the local informed consent with minimal changes from the template; and the investigator's CV.

Adverse events and protocol deviations are infrequent; however the study team and Wake Forest IRB developed a Safety Event Reporting Form to capture unanticipated problems, protocol violations or other safety events.

The majority of SPRINT ASK clinics had their own local IRB. Most of these clinics had to complete some type of review of the protocol and/or informed consent documents. There was great variety from the institutions as to whether this local review occurred prior to the central review or after. Clinics without a local IRB had to provide proof of current certification for human subjects training.

The Central IRB system can be an efficient and effective way to manage participant safety and decrease burden on clinic staff; however it is important to recognize the burden on both the central staff and the central IRB. This presentation will elaborate on the lessons learned from the SPRINT ASK central IRB experience and discuss the advantages and limitations observed.

Contributors

James Moore
Wake Forest University Health Sciences

Jennifer Walker
Wake Forest University Health Sciences

David Reboussin
Wake Forest University Health Sciences

A44

STATTAG FOR CONNECTING R, SAS, AND STATA TO WORD: A PRACTICAL APPROACH TO REPRODUCIBILITY

ABIGAIL BALDRIDGE

NORTHWESTERN UNIVERSITY FEINBERG SCHOOL OF MEDICINE

Background

Reproducibility, wherein data analysis and documentation is sufficient so that results can be recomputed or verified, is an increasingly important component of statistical practice and publication of scientific studies. Clinical trials in particular have been subject to increased scrutiny and requirements for data sharing. To address this challenge, we have created StatTag, a free, open-source program that embeds statistical results from R, SAS, or Stata directly in Microsoft Word.

Methods

StatTag is available as a Word plugin (Windows) or standalone application (Mac) that links statistical code files to Word documents. From Word, users attach one or more code files to an active document, and use the StatTag interface to "tag" selected statistical output – estimates, tables, or figures. The user instructs StatTag to insert the selected statistical output into the Word document, whereupon StatTag invokes the appropriate statistical software and places the result within the document text. Inserted results can then after be updated automatically or on demand. All inserted results can be double-clicked to view the exact code which generated them, and will retain their linkage to the code even when the document changes hands, is redlined, or the text is copied and pasted elsewhere. The StatTag interface also allows direct user interaction with the code file; users may view, edit and re-run statistical code directly from Word. StatTag improves over other similar software in that it functions directly from Word, and it allows the usage of more than one statistical software and code.

Results

StatTag is well suited for clinical trials, where ongoing monitoring and reporting are necessary, and often require a substantial burden of time to create and maintain. Consider a CONSORT flow diagram or Data Safety Monitoring report, both of which are rich in data and may be regularly updated over an enrollment period of months or years. By using StatTag, the analyst, investigator, or coordinator can develop the flow chart or report in Word and populate the results using StatTag, allowing the results to be updated anytime the enrollment changes, without the effort and potential risk of human transcription errors.

StatTag has been developed for Windows and Mac platforms, and for the most common statistical programs; R, SAS and Stata. StatTag has been downloaded by over 450 users, including many in the non-exclusive fields of behavioral science (11%), biostatistics (34%), epidemiology (26%), clinical medicine (21%) and public health (24%).

Discussion

StatTag will improve the efficiency and quality of trial monitoring and publication, will be essential for conduct of efficient reproducible research, and will be broadly useful to all involved in clinical trials research.

Contributors

Leah Welty
Northwestern University Feinberg School of Medicine

Luke Rasmussen
Northwestern University Feinberg School of Medicine

Eric Whitley
Northwestern University Feinberg School of Medicine

A45

DYNAMIC MANUSCRIPT DEVELOPMENT AND TRACKING OF SCIENTIFIC PRODUCTIVITY AND IMPACT

ASHLEY HOGAN

GEORGE WASHINGTON UNIVERSITY

Under pressure from an ethical mandate and recently implemented NIH policies on dissemination of clinical trial results within 1 year of final data collection, timeliness is critical in manuscript development. The demand for pragmatic and collaborative processes is imminent to resolve the logistical and communication challenges in the development and tracking of publications. In an increasingly competitive environment, amplified and sustainable productivity cannot be understated to maximize potential public health impact.

Detailed monitoring ensures that the group remains productive while upholding the highest quality publications and presentations. Our team developed an integrated system to capture the workflow from proposal review, through analyses, writing, manuscript reviews, submission, and finally public dissemination through social media. With this system, we are able to track each stage of the manuscript process and incorporate checkpoints to ensure manuscripts meet quality standards and reproducibility through timely reviews by the writing group, Publications and Steering Committees. Using a web-based enterprise portal software, a template-based website facilitates volunteer solicitation, member identification, reviews through message boards, journal entries for agendas and minutes, and document repositories with various levels of role-based permissions to ensure security of confidential data and drafts. The customized website templates have allowed the quick deployment and sustainable management of simultaneous writing group activities. At present, over 2,000 users are working on more than 300 writing group sites among six research projects . The system also includes a downloadable citation library, citation tracking, and searchable registry of all publications and presentations with links to PubMed and full text.

The purpose of this presentation is to share the various features of the system and how each component has the potential to foster scientific integrity, increased productivity, and timely dissemination. Attendees will learn how the integration of a seamless workflow for manuscript tracking can support an enduring influence on public health research.

Contributors

Alla Sapozhnikova
George Washington University

Sharon Edelstein
George Washington University

Marinella Temprosa
George Washington University

A46

RANDOMIZED CLINICAL TRIAL ABSTRACTS SUBMITTED TO MAJOR MEDICAL MEETINGS: DO THEY INCLUDE TRIAL REGISTRATION NUMBERS AND CONFORM TO CONSORT GUIDELINES?

ELIZABETH WRIGHT

NIDDK/NIH

Introduction: Approximately 25% of randomized clinical trials (RCTs) presented at medical meetings are not published as full articles. Abstracts may, therefore, be the only source of information for these trials and should conform to consort guidelines for abstracts (Hopewell et al. PLoS Med. 2008 Jan 22;5(1):e20).

Background: A recent review of abstracts submitted to the 2015 American Society of Clinical

Oncology (ASCO) annual meeting found that 64% included the word 'randomized' in the title, 46% included the number randomized in each group, 7% included the funding source, and 93% included a trial registration number (Liu et al. JAMA Oncol. 2017 Mar 1;3(3):414-416). We sought to determine whether other meetings had similar rates.

Methods: We obtained abstracts for three 2017 annual meetings: the American Association for the Study of Liver Disease (AASLD), the American Diabetes Association (ADA), and Kidney Disease Week (KDW). We reviewed abstracts containing the word randomized (randomised) and selected those reporting results from RCTs. Abstracts were classified as primary if they appeared to be reporting the primary outcome of the trial. We then reviewed the primary abstracts using consort guidelines. We searched ClinicalTrials.gov and the World Health Organization International Clinical Trials Registry Platform for registration numbers if not included. Late breaking abstracts or print only abstracts were not included in this analysis.

Results: The word "randomized" appeared in 134 (6%) of 2239 AASLD abstracts, 305 (14%) of 2234 ADA abstracts, and 210 (6%) of 3696 KDW abstracts. We excluded 131 of the 649 abstracts, 36 because they were animal studies, 48 because the word "randomized" was used only in the background or conclusion sections, 32 because they were meta-analyses or systematic reviews, and 15 for other reasons. 313 of the 518 reports of RCTs were classified as primary (60%). This percent was highest for ADA (186/271=69%) compared to AASLD (49/96=51%) and KDW (78/151=52%).

Review of the 313 primary abstracts found that 32% (53% AASLD, 20% ADA, 46% KDW) included the word "randomized" in the title, 94% included the number randomized, but only 61% (65%, 59%, 62%) included the number randomized in each group, 56% (0%, 64%, 73%) included the funding source, and 15% (35%, 10%, 13%) included trial registration numbers. The median number randomized was 60. Registration numbers were included for 10% of trials with <60 randomized compared to 19% of trials with 60 or more randomized. Registration numbers were found for 120 (76%) of the 157 abstracts with 60 or more randomized, 107 were registered at ClinicalTrials.gov.

Conclusions: Most trials were registered but only 15% of primary abstracts included registration numbers. Including registration numbers in abstracts would enable linkage of all trial publications for systematic reviews and other analyses (Altman et al. Trials. 2014 Sep 23;15:369). None of these 3 societies mentioned trial registration in their instructions for abstract submission. In contrast, ASCO, which had a 93% registration rate, asks for trial registration number during the abstract submission process. Other societies should consider doing the same.

Contributors

A47

THE NATIONAL DRUG ABUSE TREATMENT CLINICAL TRIALS NETWORK (NDAT CTN) STRATEGIC PLANNING GROUP ON RETENTION: LESSONS LEARNED AND BEST PRACTICES

MITRA LEWIS

THE EMMES CORPORATION

Introduction/Background: Despite its fundamental importance, retention remains among the most challenging aspects in the successful implementation of randomized controlled trials. Barriers to achieving high retention rates may be more pronounced in CTN trials, where participants generally have a current or previous substance use disorder (SUD).

The Center for Clinical Trials Network (CCTN) manages NIDA's NDAT CTN and works with a Clinical Coordinating Center (CCC), a Data and Statistics Center (both housed at The Emmes Corporation), and 13 research Nodes throughout the United States to conduct multisite substance use treatment studies.

Objective: To disseminate the strategic plan for retention in CTN clinical trials as assembled and distilled from the extensive knowledge of the NDAT CTN Nodes and Clinical Coordinating Center.

Methods: In July 2017, the CCC hosted the NDAT CTN Recruitment and Retention Strategic Planning Meeting at Emmes Corporate headquarters. This daylong meeting brought together a multi-disciplinary focus group with vast experience in clinical trial research with the SUD population. Meeting participants shared tools, media, materials, and experiences across regions and trials, synthesizing them to develop recommendations for general best practices for future studies.

Results/Conclusions: Effective retention must be carefully considered both in the protocol development phase, including budget development, and throughout trial implementation. Efficacious planning prior to study initiation must include an evaluation of the following: eligibility criteria, site-specific barriers, and population-specific barriers including those associated with gender and ethnic/racial minorities. Each inclusion and exclusion criterion should be scrutinized prior to protocol finalization to determine their potential impact on enrollment. Population-specific barriers such as desirability of study interventions/procedures, population comorbid health, psychiatric and social issues, and involvement in the criminal justice system should be considered. Site-specific barriers, including competing research priorities, limited clinical trial experience, and lack of site "buy in", should inform the site selection processes as well as training and other pre-implementation activities.

Procedures for increasing retention at the site level and monitoring overall progress throughout trial implementation are equally tantamount. Sites will benefit from developing active retention and follow-up plans customized to their participant demographics. Further, teams should develop tracking systems to monitor both retention efforts to meet enrollment goals. For multi-site trials, centralized reports such as those generated on the NDAT CTN website are beneficial. This website includes standardized Trial Progress Reports (TPRs) and Data Status Reports (DSRs) that allow study teams to track retention status from the date of first participant first visit to last participant

last visit; additionally, reports may be customized for retention monitoring by summarizing attendance at follow up visits or summarizing participant disposition (e.g., randomized, in active treatment, in follow up).

Overall, planning retention efforts and approaches during study development affords research staff the necessary tools that each site needs to perform retention efforts throughout trial implementation. The strategic planning group delineated a variety of potential strategies and techniques that can be applied for the successful retention of participants in clinical trials.

Contributors

Dagmar Salazar
The Emmes Corporation

John Farley
San Francisco Department of Public Health

Eve Jelstrom
The Emmes Corporation

Frankie Kropp
University of Cincinnati College of Medicine

Tim Matheson
San Francisco Department of Public Health

A48

ANALYSIS METHODS FOR INDIVIDUALLY RANDOMISED TRIALS WITH PARTIAL CLUSTERING OF OUTCOMES

JANE CANDLISH

UNIVERSITY OF SHEFFIELD

Background: In an individually randomised trial we might expect that interventions delivered in groups or by the same health professional result in clustering of outcomes for participants treated in the same group or by the same health professional. A partially nested trial is a design where the clustering only occurs in one arm of the trial, commonly the intervention arm. This is further complicated in adaptive or staged interventions which often produce more complex forms of clustered data. In these trials intervention stages are delivered based on individual responses to previous stages, one or more of the intervention stages can lead to different levels of clustered outcomes. Where possible, it is important to measure and account for the between-cluster variability in trial design and statistical analysis.

Objective: This study compares several statistical models to analyse individually randomised trials with a continuous outcome and differential clustering of outcomes, for both partially nested trials and adaptive intervention trials with nested outcomes.

Methods: We perform a series of simulation studies assessing the performance of different statistical analysis models. The impact of cluster sizes, data coding of the non-clustered participants, intracluster correlations (ICC), and differential variance between trial arms on statistical inference will be presented. Analysis models include: simple linear regression with different variance estimation methods (standard variance estimator for ordinary least-squares regression, bootstrap estimators and cluster robust estimators), mixed effects models, and partially nested mixed effects models.

Results: All models provided unbiased intervention effect for the partially nested trials. For partially nested trials the heteroscedastic partially nested mixed-effects model (using Satterthwaite degrees of freedom) maintained type I error rates (mean 0.044, SD 0.014) and provided unbiased ICC estimation. However, with very few clusters and small cluster sizes, the partially nested models did not control the type I error rates. For adaptive interventions a suitable mixed effects model could not be parameterised without causing bias of the overall intervention effect. However, with ICC = 0.05 the naive simple linear regression model resulted in a small inflation of the mean type I error rate of 0.057 (SD 0.010).

Conclusion: For partially nested trials we recommend the use of a heteroscedastic partially nested mixed-effects model, which models the clustering in only one arm. For adaptive interventions with nested outcomes, it was not possible to account for clustering of the outcomes imposed by the design using mixed effects models. Therefore, we caution the use of such trial designs when the ICC is expected to be large. It has been previously recommended that treatment-related clustering is accounted for in trials. Though this reflects the design of the trial, we recommend trialists should

not adjust for clustering which is not present for all in the intervention arm using a mixed effects model and when this clustering is based on a post-randomisation factor.

Contributors

M. Teare
University of Sheffield

Judith Cohen
University of Sheffield

A49

PSYCOMETRIC AND INITIAL CLINICAL VALIDATION OF THE BLUEBELLE WOUND HEALING QUESTIONNAIRE (WHQ): A NEW OUTCOME MEASURE FOR POST-DISCHARGE ASSESSMENT OF WOUNDS FOR SURGICAL SITE INFECTION

RHIANNON MACEFIELD

UNIVERSITY OF BRISTOL

Background

Surgical site infections (SSIs) are problematic for patients and have high associated healthcare costs. Reduction of SSI is therefore high priority for patients, for health services and for research. Measuring rates of SSI is critical in trials evaluating interventions to minimise SSI, however, accurate assessment is challenging because many SSIs occur after the patient has left hospital. Face-to-face follow-up assessment is resource intensive and costly and well-developed post-discharge questionnaires for patient-report are lacking. Existing tools for assessing SSI include the U.S. Centres for Disease Control and Prevention (CDC) criteria and ASEPSIS grading scale however their applications for use after hospital discharge are limited, they have not been developed for patient completion or been formally validated. There is a need for better and more efficient methods for post-discharge SSI outcome assessment and trial data collection using patient-centred tools. The Bluebelle WHQ was developed specifically to address this need [1]. This study examined the acceptability, reliability and validity of the WHQ in a large sample of patients undergoing abdominal surgery.

Methods

Data were collected within a purpose-designed cohort study and a pilot RCT. Patients completed the WHQ (self-assessment) and debriefing questions to assess WHQ acceptability and ease of completion approximately 30 days after surgery. Questionnaires were administered by post. A sub-sample (n=50) were asked to complete the WHQ twice within a few days (test-retest reliability). HCPs completed the WHQ (observer-assessment) approximately 4-8 weeks after surgery by telephone or during clinic visits. Face-to-face reference SSI diagnoses were made by an independent assessor using the CDC criteria. Response rates, missing data and patient/observer responses were examined. Factor analyses explored the dimensionality of data (WHQ internal structure) and a suitable scoring system. Sensitivity and specificity of WHQ self-assessments for discriminating SSI/no SSI against the reference CDC diagnosis were examined.

Results

Some 591/792 (70.8%) WHQ self-assessments and 597/791 (74.4%) observer-assessments were completed. Test-retest reliability was good (kappa < 0.6 for the majority of items). Agreement in responses between self- and observer-assessments was high overall, although patients showed a trend to report signs and symptoms slightly more severely and low fidelity was observed for some of the rarer wound care interventions. Data supported a single scale (unidimensional construct) with strong internal consistency (Cronbach's alpha <0.8). Sensitivity and specificity for

discriminating SSI/no SSI was high (area under ROC curve=0.9056). High levels of missing data in some sub-items meant revisions to the questionnaire format was required.

Conclusions

The WHQ demonstrates good reliability and validity and it is suitable for post-discharge assessment of wounds by patients or observers. The revised version now needs further validation and performance testing in other patient populations and surgical specialties. The feasibility of a patient-centred method for collecting self-taken wound photographs to supplement the WHQ is currently being explored.

[1] Macefield RC, Reeves BC, Milne, TK, Nicholson A, Blencowe NS, Calvert M, et al. Development of a single, practical measure of surgical site infection (SSI) for patient report or observer completion. Journal of Infection Prevention. 2017;18(4):170-179

Contributors

A50

DEVELOPMENT OF A CORE OUTCOME SET CAPTURING KEY CONCEPTS RELEVANT TO SAFE AND EFFICIENT EVALUATION OF INNOVATIVE INVASIVE PROCEDURES

KERRY AVERY

UNIVERSITY OF BRISTOL

Background:

Methods for introducing invasive procedures (IP: surgical procedures with/out a device) into clinical practice are haphazard and largely ungoverned, with few RCTs and fewer well designed and conducted early phase studies. This contrasts the highly regulated research environment for introducing new pharmaceutical products. Unregulated introduction of IPs has resulted in several scandals, such as the recall of counterfeit spinal fusion screws in the US, removal of faulty silicone breast implants in Europe and ongoing concerns over the use of vaginal mesh implants in the UK, following association with previously unrecognised adverse events (AEs). While systems exist for reporting outcomes of some established IPs (e.g. American Joint Replacement Registry, UK National Adult Cardiac Surgery Audit, UK National Oesophago-Gastric Cancer Audit), outcome definition, selection, measurement and reporting in early phase studies of innovative IPs lacks standardisation. This prevents data syntheses from independent studies and delays identification of emerging problems. Effective systems and mandatory requirements for surgeons to report outcomes when developing/implementing new IPs are also lacking, meaning AEs are likely under-reported.

Development of a core outcome set (COS), a minimum mandated set of outcomes to be measured and reported in all early phase studies of innovative IPs and accompanied by transparent reporting guidelines mandated via regulatory agencies and publications, is necessary to promote uniform and evidence-based assessment so that innovative IPs are introduced safely and efficiently.

Aim:

To describe methods and early results (steps 1-2 below) of the identification of outcome domains for minimal reporting of innovative IPs.

Methods:

1. Identify existing systems for reporting outcomes of innovative IPs;

2. Categorise these outcomes (e.g. AEs, complications, symptoms, function) into ‘conceptual’ domains;

3. Apply the conceptualised domains within case studies of innovative IPs to refine them and establish how current AE reporting maps onto domains;

4. Iteratively engage key stakeholders to consider the above (e.g. surgeon innovators, patients, device manufacturers, regulators, journal editors) and develop guidelines for reporting of innovative IPs.

Results:

Examples of outcome reporting systems identified include the US Food and Drug Administration Device Regulatory Strategy, Clavien-Dindo classification of surgical complications, UK NICE Interventional Procedures Guidance and Medicines and Healthcare products Regulatory Agency). No single system provides detailed mandated guidance on outcome domains relevant to comprehensively evaluate innovative IPs specifically. Early findings indicate that a broader conceptualisation of outcomes, beyond traditional outcomes such as AEs and complications, is needed to begin to tackle the issues and provide transparent data.

Conceptualised domains identified so far include:

(i) Innovation delivered with intended effect;

(ii) Innovation delivered with unintended effect;

(iii) Innovation delivered and caused anticipated AE;

(iv) Innovation delivered and caused unanticipated AE;

(v) Innovation abandoned (intraoperative/postoperatively);

(vi) Innovation associated with longer-term unintended/unanticipated effects/AE.

Conclusions:

There is an urgent need to develop and implement better methods for early phase evaluation of innovative IPs. Identifying core outcome domains to inform selection of specific agreed outcomes for IPs is the first necessary step in this process. Further work is needed to test these conceptualised domains in specific innovations and with key stakeholders.

Contributors

Shelley Potter
University of Bristol

Nicholas Wilson
University of Bristol

Rob Hinchliffe
University of Bristol

Sian Cousins
University of Bristol

Angus McNair
University of Bristol

A51

"TREATMENT AS USUAL" AS THE COMPARATOR IN RANDOMISED CONTROLLED TRIALS: A SCOPING REVIEW OF THE USE OF THE TERM OVER THE PAST FIVE YEARS IN SIX INTERNATIONAL GENERAL MEDICAL TRIAL JOURNALS.

KIRSTY LOUDON

UNIVERSITY OF STIRLING

Background:

Clinicians and policymakers eager to implement Evidence Based Practice often favour Randomised Controlled Trials (RCTs) that compare the new intervention(s) to be tested in clinical settings with usual care. In particular, they favour pragmatic trials because they are designed to determine whether or not an intervention works in "usual care" settings. The comparator 'usual care', however, is often poorly defined. Readers of the papers and reports arising from studies are then unable to determine how the new intervention would compare with usual care in their own setting, leading to uncertainty in predicting the successful implementation of the intervention. The aim of this scoping review was to investigate how "usual care" is defined and described in RCTs.

Methods

We searched MEDLINE (PubMed) using key words "usual care", "routine care", "standard practice" and "standard care" for six international general medical journals (NEJM, The Lancet, The British Medical Journal, Trials, PLOS Medicine, and JAMA) which published RCTs in 2016. One reviewer assessed the abstracts and full texts for inclusion and 20% of the screened articles were independently checked for inclusion by four reviewers. Disagreements were resolved through discussion. Seven reviewers extracted data on: the intervention description, "usual care" descriptions, use of the word "pragmatic" in title or abstract or full text and RCT settings and country. Two researchers then independently assessed the quality of the "usual care" description using three categories: poor, unclear, or high, the latter of which represented sufficient description to support judgements about applicability of the results to other settings.

Results

We screened 384 articles and 68 RCTs were included for assessment of the usual care description. There were 32(47%) RCTs with descriptions of "usual care" that were judged as being sufficiently clear to facilitate applicability judgements in other settings. There were also nine RCTs which had a comparator group of "enhanced usual care", meaning the comparator was not actually usual care. Only thirteen RCTs (19.1%) with a "usual care" comparator were described as pragmatic in the main text. The most common setting for RCTs with a usual care comparator was primary care (n = 21), with the USA the most likely country for an RCT with a usual care comparator in our sample.

Conclusions

Although 32/68 "usual care" descriptions were of a high standard and included clinical guidelines or web-links facilitating understanding, there was confusion over the terms "usual care" and

"enhanced usual care". Trial teams need to be more careful in study design and descriptions of usual care as poor description of the comparator will lead to problems using or applying the results in clinical practice.

This research should alert trial teams, designing and describing pragmatic RCTs, to ensure that "usual care" as well as the intervention is accurately described and in sufficient detail to fulfil CONSORT statement guidance (e.g. by using the TIDieR checklist), particularly the CONSORT extension for pragmatic trials. Journals could help by providing guidance to authors on how to describe usual care comparators.

Contributors

Merrick Zwarenstein
Western University

Shaun Treweek
University of Aberdeen

Ifeoma Elueze
University of Toronto

Zehra Laiwalla
University of Toronto

Natalie Knight
University of Stirling

A52

DESIGN ANALYSES OF RANDOMIZED CLINICAL TRIALS SUPPORTING FDA CANCER DRUG APPROVALS

LUDOVIC TRINQUART

BOSTON UNIVERSITY SCHOOL OF PUBLIC HEALTH

Purpose: Statistical significance continues to drive the design and interpretation of randomized controlled trials (RCT), in particular in the context of cancer drug approval. We examined the Type S error risk – claiming a new drug is beneficial when it is actually detrimental - and the exaggeration ratio - the factor by which a significant estimated effect differs from the true effect – as ways to reemphasize the direction and magnitude of effects in RCTs.

Methods: We systematically reviewed pivotal RCTs supporting Food and Drug Administration (FDA) approval of cancer drugs between 2007-2016. We extracted outcome data for overall survival (OS), progression-free survival (PFS), and response outcomes from FDA reviews. For each RCT, we estimated the type S error risk and exaggeration ratio by considering replicated RCTs of equal size and a range of true treatment effects.

Results: We analyzed 43 trials for 39 approved drugs. Across 38 RCTs reporting OS, the median type S error risk was 0.00% [Q1-Q3, 0.00%-0.01%] and 3.56% [0.40%-6.74%], for true hazard ratios of 0.7 and 0.9, respectively, meaning that one can be confident about the direction of the effect. The corresponding exaggeration ratios were 1.09 [1.01-1.11] and 1.30 [1.13-1.42], indicating that the true effect is overestimated by a median 9% and 30%. We found similar results for PFS and we found moderately larger type S error risks and exaggeration ratios for response outcomes.

Conclusion: Our findings highlight how type S error and exaggeration ratios give additional insights on the replicability of RCTs, regardless of statistical significance. Our findings also provide quantification of the winner's curse, in which pivotal RCTs supporting cancer drug approval tend to be overly optimistic.

Contributors

Emily Lord
Boston University School of Public Health

Isabelle Weir
Boston University School of Public Health

A53

CHOOSING A BALANCING CRITERION FOR COMPLEX CLUSTER-RANDOMIZED TRIALS WITH CONTINUOUS BASELINE VARIABLES

JODY CIOLINO

NORTHWESTERN UNIVERSITY

Background: Randomization algorithms that control baseline variable imbalance are ideal in cluster-randomized controlled trials (C-RCTs), but randomization algorithms for C-RCTs often carry increased complexity. Many authors recommend implementation of covariate-constrained randomization as it can efficiently control imbalance in a large number of covariates. The general algorithm is:

1. Enumerate all or a large subset (e.g., 100 thousand) of all possible allocation schemes;

2. Evaluate "imbalance" for each iteration;

3. According to a pre-defined criterion, define a subset of iterations meeting this criterion of "acceptable" (im)balance;

4. Randomly select one scheme from this subset for implementation.

The majority of literature surrounding the imbalance metric (step 2 above) assumes categorical baseline variables and two study arms. In an ongoing PCORI-funded trial (PI: Darius Tandon, PhD; Northwestern University), we employed a variation of this procedure to control imbalance in three continuous baseline variables across three arms. We chose the p-value corresponding to the Kruskal-Wallis (KW) test as our criterion for "balance"; we deemed an allocation scheme "acceptable" if the KW p-value for each of the three variables was larger than 0.30. Though utilization of this criterion resulted in adequate baseline variable balance, we used simulation studies to explore other metrics of baseline variable imbalance.

Methods: We simulated 10 thousand C-RCTs with three arms and three continuous baseline variables as follows:

1. Assume three variables for each of three trial arms come from a multivariate normal distribution with some pre-specified level of correlation: no correlation, compound symmetry structure, a single large correlation between two variables, a "realistic" structure inspired by our example trial.

2. Calculate imbalance metrics for each simulated C-RCT:

a. The smallest of three ANOVA p-values comparing mean of each variable across the three arms

b. The smallest of the KW p-values (as in our trial)

c. MANOVA p-value simultaneously comparing means across arms

d. The smallest of nine (all possible pairwise comparisons) independent t-test p-values

e. The smallest of nine Wilcoxon Rank-Sum (WRS) test p-values

We simulated under both balanced and intentionally imbalanced conditions.

Results: The correlation structure assumptions had minimal impact on overall findings, and the t-test criterion was uniformly the most conservative (48% of the time), followed by the WRS (29% of the time), ANOVA (21%), MANOVA (3%), and the KW test (<1%). If we deemed p-value > 0.30 as acceptable for each metric, the KW, t-test, and WRS criteria detected imbalance in 100% of the simulated imbalanced scenarios; however, the ANOVA and MANOVA criteria failed to detect this imbalance just 0.03% and 0.34% of the time in these scenarios.

Conclusions: Our criterion based on the KW test p-value > 0.30 to signify "acceptable" balance was not the most conservative of those explored, but it appropriately identified imbalance in 100% of the intentionally imbalanced scenarios. Since all explored metrics are related, results suggest modified constrained randomization algorithms involving any of these imbalance metrics for continuous baseline variables will ensure robust simultaneous control over multiple continuous baseline variables across multiple arms.

Contributors

Alicia Diebold
Northwestern University

Jessica Jensen
Northwestern University

Kimberly Koloms
Northwestern University

Darius Tandon
Northwestern University

A54

THE CHALLENGES AND OPPORTUNITIES OF MOBILE TECHNOLOGY: A QUALITATIVE EXAMINATION OF INVESTIGATORS' EXPERIENCES AND RECOMMENDATIONS

WILLIAM WOOD

UNIVERSITY OF NORTH CAROLINA SCHOOL OF MEDICINE

INTRODUCTION: Sensors, smartphone apps, and other mobile technology are increasingly being used to collect objective trial endpoint data, creating opportunities for the development of novel endpoints, enhanced patient experiences, and improvements in the quality and efficiency of clinical trials. This new technology, however, also creates new challenges for investigators and site personnel—challenges that are critical to understand and address if we are to realize the promise mobile technology holds.

METHODS: To understand investigator experiences in trials that use mobile technology for data collection, the Clinical Trials Transformation Initiative (CTTI) conducted a series of thirteen one-on-one, semi-structured telephone interviews with investigators at both academic and independent research sites. Interviews were approximately one and a half hours. All interviewed investigators had conducted both traditional and mobile trials.

RESULTS: Interview findings reveal investigator perspectives on the advantages and disadvantages of mobile clinical trials for investigators, comparisons of traditional vs. mobile clinical research, training and budget considerations, and guidance for other investigators who are considering conducting mobile trials.

CONCLUSIONS: Mobile technology offers many advantages to both investigators and patients—from increased data quality, to better monitoring and reduced patient burden—but also presents new challenges, including new costs, training requirements, and staffing needs. These findings provide important, practical insights from investigators with hands-on experience that will help the research enterprise better plan and conduct future clinical trials using mobile technology.

[Please note that this work, if selected, would be presented by Dr. Wood on behalf of the CTTI MCT Stakeholder Perceptions Project Team (https://www.ctti-clinicaltrials.org/projects/stakeholder-perceptions), who would also be available as alternates in the unlikely event of an unexpected conflict.]

Contributors

A55

MOBILE TECHNOLOGY AND CLINICAL TRIALS: PATIENT PERSPECTIVES AND OPPORTUNITIES TO REDUCE THE BURDEN OF PARTICIPATION

VIRGINIA NIDO

GENENTECH, A MEMBER OF THE ROCHE GROUP

INTRODUCTION: The growing use of mobile technology to collect objective trial data holds tremendous potential to minimize the burden of participation on patients and increase accessibility of clinical trials to underrepresented populations. As such, the Clinical Trials Transformation Initiative (CTTI), a public-private partnership co-founded by Duke University and the FDA, convened a multi-stakeholder project team to examine both the opportunities and potential barriers to implementation.

METHODS: An online survey was conducted to understand patient perspectives of mobile technology commonly used in clinical trials, as well as preferences for mobile vs. traditional clinical trials and associated considerations. The survey was developed with input from research sponsors conducting such trials, patients and advocates, investigators, and other relevant stakeholders, and underwent cognitive interview testing with patients in the four targeted disease/therapeutic areas (arthritis, cardiovascular disease, Parkinson's disease, and diabetes).

RESULTS: 193 respondents provided analyzable responses (mean age of 60 years, 62% female, 88% white / non-Hispanic). Strong preferences were observed for trials using mobile technology as compared to traditional trials, and preferred characteristics of mobile trials and mobile technology were identified, including device attributes, privacy considerations, and communication and visit preferences.

CONCLUSIONS: Although the use of mobile technology presents new challenges that must be addressed, our findings suggest opportunities to meet patient needs and enhance participation in ways that traditional trials have not been able to accomplish, and identify important considerations for researchers who are conducting such trials.

[Please note that this work, if selected, would be presented on behalf of the CTTI MCT Stakeholder Perceptions Project Team (https://www.ctti-clinicaltrials.org/projects/stakeholder-perceptions), who would also be available as alternates in the unlikely event of an unexpected conflict.]

Contributors

A56

IDENTIFICATION OF HIGH-EFFICACY SUBGROUPS IN CLINICAL TRIALS

GARRICK WALLSTROM

STATISTICS & DATA CORPORATION

Purpose: In many studies, we seek to identify a subgroup of patients in which an active treatment is most effective compared to a control. Existing methods for finding such high-efficacy (HE) patient subgroups often find overly-complex subgroups and do not control the type 1 error rate, which can lead to low replication rates in future trials. This study developed and tested a novel strategy that identifies HE subgroups while controlling the type 1 error rate.

Methods: Our novel approach conducts a constrained Monte Carlo search using a set of select baseline characteristics to identify an HE subgroup. The search is constrained to enforce an interpretable subgroup definition, limit the number of characteristics used to define the subgroup, and ensure adequate subgroup size. We use permutation tests to accurately assess statistical significance of treatment differences in the subgroup while accounting for multiple testing. To evaluate our method, we simulated typical data for a clinical trial comparison of active and control groups with 60 subjects per group. The primary endpoint was the change from baseline in an endpoint on a 0-4 scale. Treatment groups were compared using a t-test. We sought HE subgroups defined using up to two out of three simulated baseline measures. We examined two scenarios. In Scenario 1, there were no true HE subgroups, so any statistically significant HE subgroups found were false positives. In Scenario 2, we simulated a true HE subgroup of subjects that represented ~60% of all subjects. For Scenario 2, we calculated the positive predictive value (PPV), the percentage of the found subgroup that is in the true HE subgroup, and sensitivity, the percentage of the true HE subgroup that is in the found subgroup. For each Scenario, we applied our method to 500 repetitions and calculated the statistical significance of the identified subgroup using the unadjusted t-test and the permutation test. Additional testing scenarios were also examined to explore the impact of the number of baseline characteristics and search constraints on performance. All tests were conducted using a 2-sided significance level of 0.1.

Results: In Scenario 1, the false positive rate was 78% using unadjusted t-tests and 10% using permutation testing, the latter of which is consistent with the 0.10 significance level. In Scenario 2, the subgroups found had an average PPV of 90% (SD=17%) and an average sensitivity of 82% (SD=17%). Using the unadjusted t-test, 100% of the subgroups were statistically significant compared to 91% using the permutation test.

Conclusions: We have developed a powerful approach for identifying HE subgroups while controlling the type 1 error rate. Our approach successfully found true HE subgroups with 90% PPV and 82% sensitivity. Further, our approach successfully limited the rate of false findings, in stark contrast to the use of unadjusted t-tests which had a false positive rate of over 75%. This approach will aid in post-hoc analysis of clinical trials toward efficient design and patient enrichment of future trials as well as targeted patient subgroup manuscript discussions.

Contributors

Kirk Bateman
Statistics & Data Corporation

Dale Usner
Statistics & Data Corporation

A57

MILESTONE PREDICTION FOR TIME-TO-EVENT ENDPOINT MONITORING IN CLINICAL TRIALS

FANG-SHU OU

MAYO CLINIC

Predicting the times of milestone events for interim and final analysis in clinical trials helps resource planning. The goal of this project is to develop robust algorithms to aid milestone prediction for time-to-event endpoints.

We investigate several easily implemented methods, in both frequentist and Bayesian frameworks, for predicting when a milestone event is achieved. We show that it is beneficial to combine multiple prediction models to craft a better predictor via prediction synthesis. Furthermore, a Bayesian approach provides a better measure of the uncertainty involved in the prediction of milestone events. We compare the methods through two simulations; one where the model has been correctly specified and one where the models are a mixture of 3 incorrectly specified model classes. We then apply the method on a real clinical trial data, NCCTG N0147. The performance using Bayesian prediction synthesis is very satisfactory, i.e. the predictions are within 20 days of the actual milestone (interim at 50% of event) time after 20% of events were observed.

In summary, the Bayesian prediction synthesis methods automatically perform well even when the model is incorrectly specified or data collection is far from homogeneous. We are developing an R shiny app to carry out the aforementioned prediction in a user friendly fashion.

Contributors

Martin Heller

Qian Shi
Mayo Clinic

A58

EXPANDING ON ACCEPTED WISDOMS OF STATISTICAL DESIGN AND ANALYSIS - STRATIFIED RANDOMIZED CONTROLLED TRIALS

NICOLA GREENLAW

UNIVERSITY OF GLASGOW

Introduction

As students of medical statistics, and as early career researchers, we were provided with various ”rules of thumb” or ”accepted wisdoms”, many of which have become imprinted in our memories. Throughout our careers as clinical trial statisticians, we have designed and analyzed many randomized trials using these general principles, without necessarily questioning whether what we were doing was optimal.

Background

There exists a general understanding that if an association between a baseline covariate and the primary outcome is expected in a randomized clinical trial, this should be accounted for via stratification methods in both the design and analysis(1). Using an example of a typical randomized clinical trial, it has been proven by simulations that there can be a difference in the level of significance observed and a reduction in power if a study design incorporating the stratification is not analyzed accordingly(2).

For example, when studying an outcome that is expected to vary with age, one would often stratify the randomization scheme, randomizing younger and older participants separately, and then adjust the final analysis for these age groups. However, upon a brief literature search, there was no evidence as to whether there would be any additional benefit obtained by incorporating such a stratification variable as a continuous covariate, rather than the binary factor used for stratification. Furthermore, many clinical trials stratify on multiple baseline covariates (e.g. age and gender), and adjust for each stratification variable in an additive model; however, it may be worth contemplating whether even more power could be gained by permitting the model to allow for an interaction between stratification variables, as well as the main effects.

Research Questions

Given a single continuous baseline covariate, dichotomized for use as a stratification variable, is there more power to be gained from fitting a model adjusted for the continuous value rather than the stratification value? With the addition of a truly binary stratification factor, what is the impact of adjusting for an additional interaction term?

Methods

We shall explore via simulations, some scenarios to determine which methods permit the most power and whether this depends on the situation.

Discussion

Clinical trial statisticians are required to apply robust statistical methodology. In a busy working environment, this is often achieved by following general principles. Our aim is to question some of these accepted wisdoms and investigate whether they can be relied upon.

References:

(1) ICH E9 Expert Working Group. Statistical Principles for Clinical Trials: ICH Harmonized Tripartite Guideline. Statistics in Medicine 1999; 18:1905 – 1942.

(2) Kahan BC, Morris TP. Improper analysis of trials randomised using stratified blocks or minimisation. Statistics in Medicine 2012; 31: 328-340. DOI: 10.1002/sim.4431

Contributors

Alex McConnachie
Robertson Centre for Biostatistics, Institute of Health & Wellbeing, University of Glasgow

A59

CENTRALIZED SOURCE DOCUMENT REVIEW: AN EFFECTIVE TOOL FOR REMOTE REVIEW OF DIGITIZED PAPER SOURCE DOCUMENTS

RACHEL WEBER

PENNSYLVANIA STATE UNIVERSITY COLLEGE OF MEDICINE

Background: As the Data Coordinating Center (DCC) for several federally-funded research networks, the Penn State College of Medicine's Department of Public Health Sciences plays a central role in the coordination and execution of large/multicenter clinical trials. While data collection is becoming increasingly paperless, paper source documents such as laboratory reports, consent forms, paper-based questionnaires, and drug/device labels are often utilized. These paper materials must be reviewed to verify accurate data capture.

Method: The Penn State College of Medicine's Department of Public Health Sciences DCC designed and implemented a Clinical Trial Application (CTA) that accommodates both paperless data capture and centralized monitoring of unavoidable paper source documents. For circumstances where paper source documentation exists, the documents are digitized and uploaded to the related electronic case report form within the CTA. The clinical staff are given a 7-day grace period to finalize their study data before it becomes available for DCC review. Data Management staff access the CTA to search for forms with "Unreviewed" source documentation and assess the accuracy of the uploaded files according to protocol-specific source document review guidelines. These guidelines ensure accurate data entry, document completeness, legibility, accurate participant identification, and redaction of unnecessary PHI/PII. If no inaccuracies are found, the status of the data is updated to "Reviewed without Errors." However, if inaccuracies are identified, the data is marked as "Reviewed with Errors" and a query is sent to the clinical site. The site then makes the appropriate correction(s) and responds to the query. Finally, Data Management staff review any corrections made by the site, resolve the query, and mark the data as "Reviewed without Errors."

Results: This review process is typically accomplished within 6 calendar days of the site's grace period ending. Having the ability to perform these reviews remotely allows the DCC to identify and address discrepancies in data entry, inconsistencies in reporting laboratory results, and consenting errors.

Conclusion: Centralized review of digitized paper source documents saves time and money while supporting data integrity. Timely review is especially important for consent verification, primary endpoint data, and identifying errors early on in a site's study participation. It also provides significant cost savings by reducing or eliminating the need to travel to each clinical site to complete source document review.

Contributors

Beth Gern
Pennsylvania State University College of Medicine

Aimee Merchlinski
Pennsylvania State University College of Medicine

Daniel Tekely
Pennsylvania State University College of Medicine

Jennifer Lucier

A60

A CONSENSUS-DRIVEN REPORTING GUIDANCE FOR ADAPTIVE CLINICAL TRIALS: AN ADAPTIVE DESIGNS CONSORT EXTENSION (ACE)

MUNYA DIMAIRO

UNIVERSITY OF SHEFFIELD

Introduction:

Adaptive designs (ADs) can enhance trial efficiency by allowing pre-planned opportunities to use accumulating trial data to modify aspects of an ongoing trial while maintaining trial validity and integrity; however, they are not widely applied due to a number of obstacles (1–6). These obstacles include lack of practical knowledge; limited access to well-reported case studies to learn from; concerns about credibility and the potential introduction of bias. Adequate reporting is one of the leading facilitators to address some of the obstacles (1). There are deficiencies in the reporting of ADs that may diminish their credibility and limit their ability to inform future related research (7–9).

There is no existing official CONSORT guidance for ADs. Therefore, we developed a consensus-driven Adaptive designs CONSORT Extension (ACE) guideline for randomised controlled trials (RCTs) to address reporting deficiencies of ADs and mitigate some of the obstacles to the routine use of adaptive clinical trials.

Methods:

We formed a multidisciplinary, cross-sector and international consortium of key stakeholders in clinical trials research to lead the ACE guideline development process and generate a list of potential reporting items. We surveyed international key stakeholders on their perceptions about the importance of the proposed reporting items during two rounds of the Delphi process. We then convened a consensus meeting attended by 27 international participants to make final decisions.

Results:

Response rates relative to registered participants in round one and two, and across rounds of the Delphi surveys were 94/143 (66%), 114/156 (73%), and 79/143 (55%) respectively. Responders were from 21 countries and had diverse adaptive designs experiences and roles in clinical trials research.

Characteristics of registered participants and responders were very similar across rounds. Perceptions about the importance of reporting items were consistent regardless of the responders' employment sector, regulatory assessment experience, self-selected stakeholder group, and whether the primary role was a statistician.

The new items in the ACE guidance include details of the planned adaptive features, planned decision-making criteria to inform trial adaptation, trial adaptation decisions made, measures to

maintain the confidentiality of interim results, statistical properties of the design and methods used to derive statistical inference.

Discussion:

We hope that the ACE guideline will promote the use of ADs by enhancing their credibility and helping to improve their reproducibility and replicability. More so, the guideline may indirectly improve the design and conduct of ADs. Though the ACE guideline has been developed focusing on RCTs, the principles might be applicable to other ADs that are not RCTs as well.

References:

1. Dimairo M et al. Trials. 2015;16(1):585.
2. Dimairo M et al. Trials. 2015;16(1):430.
3. Jaki T. Clin Trials. 2013;10(2):344-346.
4. Morgan CC et al. Ther Innov Regul Sci. 2014;48(4):473-481.
5. Love SB et al. Br J Cancer. 2017;117(3):332-339.
6. Bauer P et al. Stat Med. 2016;35(3):325-347.
7. Stevely A et al. PLoS One. 2015;10(11).
8. Bauer P et al. Biom J. 2006;48(4):493-506.
9. Hatfield I et al. Trials. 2016;17(1):150.

Contributors

Susan Todd
University of Reading

Steven Julious
University of Sheffield

Thomas Jaki
Lancaster University

James Wason
University of Cambridge and Newcastle University

Daniel Hind
University of Sheffield

A61

DESIGNING STUDY PROTOCOLS WHEN MOBILE DEVICES ARE USED FOR DATA CAPTURE: RECOMMENDATIONS FROM THE CLINICAL TRIALS TRANSFORMATION INITIATIVE

JENNIFER GOLDSACK

CLINICAL TRIALS TRANSFORMATION INITIATIVE

Background: Using mobile devices for data capture in clinical trials offers the possibility of collecting higher quality, more informative data through the conduct of trials that are simultaneously less burdensome to patients and more efficient. To realize these benefits, we must disrupt the way we currently design and conduct research. Specifically, protocols must be designed to accommodate the use of mobile devices to capture data outside of the clinic and without the supervision of a clinician. Statistical analysis plans and standard operating procedures (SOPs) must also be developed to account for the different nature and volume of data collected as well as the unique needs of managing these devices in the field.

Methods: The Clinical Trial Transformation Initiative (CTTI) convened a multi-stakeholder team comprised of sponsors, academics, regulators, technical and patient experts. Supplementing their expertise with information obtained through a series of semi-structured expert interviews and a two day expert meeting, these authors developed a suite of recommendations and resources for use by sponsors and investigators seeking to write protocols for clinical studies that rely on mobile devices for data capture.

Results: CTTI has developed solution oriented recommendations and resources to support the efforts of sponsors and investigators in protocol design across several key areas. First, CTTI has outlined considerations, protocol and technical approaches to address concerns with data collection, data attribution, data analysis and data interpretation when mobile devices are used for data capture. Second, CTTI has developed both recommendations and resources to support sponsors and investigators in aspects of data management including data integrity, study monitoring, data governance, data storage, data security and privacy. Finally, CTTI's has issued recommendations and resources to support sponsors and investigators grappling with protocol decisions including sharing data with participants in real time, monitoring both outcomes and safety signals in real time, and preparing for and handling device management--including device failure—in the field. These recommendations and resources are supported by numerous case examples.

Conclusion: There are significant benefits associated with the use of mobile devices for data capture. In order to support sponsors and investigators striving to realize these benefits, CTTI has developed a suite of recommendations and resources to support the development of study protocols, statistical analysis plans and SOPs for studies leveraging mobile devices for data capture.

Contributors

Marisa Bolognese
The LifeRaft Group

Phil Coran
Medidata

Chris Dell
Pfizer

Ray Dorsey
University of Rochester Medical Center

Cheryl Grandinetti
US Food and Drug Administration

A62

OPTIMISING RECRUITMENT IN CLINICAL TRIALS IN SURGERY: HOW DO SURGICAL TRAINEES WORKING TOGETHER ACHIEVE SUCCESS?

J ATHENE LANE

UNIVERSITY OF BRISTOL

Introduction/objective

Trials rarely recruit well or complete on time and one contributory factor is clinician engagement. This situation is often worse in surgical trials, where preferences for specific interventions are strong and research-active senior surgeons rare. Recently in the UK, Trainee Research Collaborative (TRC) networks have formed of surgeons undergoing specialist clinical training (Resident and Fellow equivalents). These TRC networks have also designed and conducted research, including several trials which have successfully delivered on time and target1,2. We aimed to understand the key elements of successful surgical trials run by these trainee networks to develop strategies to enhance clinician engagement in other clinical areas.

Background

Over 34 surgical TRCs have been established across the UK in the last five years which allow surgical trainees to join large collaborative research groups focused either geographically or by clinical speciality. These networks can potentially generate new trials, improve recruitment to ongoing trials and develop a research-active consultant (Attending Physician) workforce. TRCs have completed several surgical trials including the ROSSINI trial in 21 UK centres led by the West Midlands1. The reasons for TRC successes are unknown but understanding them is key to potential wider translation to other clinical specialities and improving clinical training in trials.

Methods

A qualitative study based on in-depth semi-structured interviews with key surgeons, trainees and clinical trials units staff conducted in 2017. Trials and linked personnel were purposefully sampled across geographical locations in the UK to include a range of surgical specialties, clinicians and TRCs. Meetings and interviews were audio-recorded and transcribed and thematic analysis utilised. Findings will be synthesised for presentation at an external stakeholders meeting in 2018. The stakeholder meeting aims to develop strategies to enhance clinicians' engagement in trials across specialties and to inform post-graduate training for clinicians.

Results

We conducted 30 interviews including; 17 trainee surgeons, 6 Clinical Trials Unit members, 5 consultant surgeons and 2 surgical research nurses. Interviewees were either directly involved or affiliated with 10 different TRCs across the UK. Preliminary findings indicate that trainee collaboratives play an important role in conducting rigorous surgical research. The "power" of these collaboratives lies in their ability to engage and bring together multi-disciplinary teams in a supportive and integrated manner. Mentoring is a key feature both within and across

collaboratives. Challenges to trainee surgeon engagement include a perceived lack of time, recognition of trainee input and authorship issues.

Conclusions

Surgical TRCs could be a novel method to increase clinician engagement in trials and enhance their academic training. An in-depth analysis of TRCs and interviews with clinical trialists has revealed some key aspects to successful TRCs conducting trials which could potentially translate to other specialities.

References

1Pinkney TD, et al. Impact of wound edge protection devices on surgical site infection after laparotomy: multicentre randomised controlled trial (ROSSINI Trial). BMJ 2013;347:f4305.

2DREAMS Trial Collaborators and West Midlands Research Collaborative. Dexamethasone versus standard treatment for postoperative nausea and vomiting in gastrointestinal surgery: randomised controlled trial (DREAMS Trial). BMJ 2017;357:j1455.

Contributors

Clare Clement
University of Bristol

Karen Coulman
University of Bristol

Natalie Blencowe
University of Bristol

Jane Blazeby
University of Bristol

Jonathan Cook
University of Oxford

A63

UNDERSTANDING STRATEGIES AND PRACTICES THAT INFLUENCE RETENTION IN CLINICAL TRIALS THROUGH AN INTERVIEW STUDY OF TRIAL STAFF IN FIVE MULTI-CENTRE TRIALS

J ATHENE LANE

UNIVERSITY OF BRISTOL

Background

There has been more focus on developing successful recruitment strategies in clinical trials than on retaining participants and collecting clinical outcome data. Evidence on factors affecting retention has mostly focused on questionnaire response rates. A survey of UK clinical trials units identified a range of strategies to enhance retention such as participant newsletters and building good relationships with participants1.

Methods

A purposive sample of five multi-centre randomised trials was selected from the UK National Institute for Health Health Technology Assessment portfolio to represent a range of clinical specialities. In-depth interviews were conducted in 2014-2015 with team members from all these trials and three additional senior trial managers to explore strategies for collecting outcome data and retaining participants. Interview data were analysed thematically using techniques of constant comparison.

Results

Nineteen members of five trials were interviewed, including chief investigators (CI, n = 4), trial managers (n = 5), academic researchers or clinical fellows (n = 5), research nurses (n= 4) and a trial administrator. Some retention strategies such as participant newsletters or reminder letters were well recognised and incorporated in trial protocols. Other strategies were implemented responsively to address retention issues and were less formalised. Interviewees highlighted the importance of fostering positive relationships with participants, for example arranging refreshments or car parking at trial appointments and telephoning unwell participants. However, these strategies took additional time which was not always recognised by funding bodies costing research nurse time. The focus on recruitment targets by national Clinical Research Networks and funders was deemed detrimental to retention which was often overlooked to ensure that the trial reached its recruitment targets. The beliefs and research experience of trial staff also affected their confidence and willingness to collect data, especially around participant withdrawals. Staff conducting follow-up sometimes came into conflict with CIs who expected follow up of everybody whilst staff used their judgement, for example if participants were unwell and stopped data collection regarding this as unacceptable. Clinicians also identified that their patient advocacy role could be conflicted by outcome data collection which they consequently stopped. Small non-monetary incentives were valued by trial staff to legitimise continued contact with participants and thus increase outcome collection.

Conclusion

Trial staff roles and underlying beliefs influence retention practices and combined with an overemphasis on recruitment targets can be detrimental to motivation and retention. There is a need to consider how to train and support staff involved in follow up and highlight the importance of retention to funding bodies and other organisations.

References

1Bower P et al. Interventions to improve recruitment and retention in clinical trials: a survey and workshop to assess current practice and future priorities. Trials, 2014. 15(1): 399.

Contributors

Anne Daykin
University of Bristol

Clare Clement
University of Bristol

Carrol Gamble
University of Liverpool

Anna Kearney
University of Liverpool

Jane Blazeby
University of Bristol

A64

OPTIMIZING PHASE II CLINICAL TRIALS

JAY SIEGEL

SOCIETY FOR CLINICAL TRIALS

BACKGROUND: Phase 2 trials in product (drug or device) development routinely are designed and powered in a sub-optimal manner.

OBJECTIVE: In this paper, I present some observations from 35 years in product development (FDA and industry) and I propose and demonstrate a rational, 4-step approach to phase 2 design.

APPROACH:

1 Strategize: Rather than design the trial around standard objectives, ie, determination of efficacy and dose selection, start by determining what are the critical questions to be answered and designs accordingly. A simple approach for that determination is presented.

2 Estimate: Selects various plausible truths (i.e., semi-quantitative answers to the critical questions) and various outcomes. Working with the expert development team, one estimates the (prior) probability of each truth, and the value associated with each potential outcome.

3 Simulate: For each truth state, run thousands of trial simulations, determining outcomes.

4 Optimize: Test various tentative post-trial decision rules (e.g., go / no-go cut points, dose selection criteria) to determine which rules optimize value (using the estimates). Then repeat simulations with varying sample sizes and select the decision rule–sample size combination that optimizes asset value.

SUMMARY: Using this approach, phase 2 trials in product development can be designed to address optimally the decisions which will follow, and can be sized to optimize the value of the asset under study.

Contributors

A65

LEVERAGING EXISTING DATABASES FOR CLINICAL TRIAL DATA

HEATHER KOPETSKIE

RHO, INC.

Introduction: There is increasing interest in leveraging electronic health record (EHR) data to conduct observational studies and clinical trials. However, such efforts are complicated due to patient privacy concerns, complexity of data extraction methods, and variations in EHR systems at different sites. The Clinical Trials in Organ Transplantation (CTOT)-20 study is being conducted across 5 North American centers to improve understanding of outcomes after lung transplantation. The study has leveraged multiple existing data sources, both at the site and through external registries, to reduce the amount of data collected through the Electronic Data Capture (EDC) system.

Data: Every solid organ transplant performed in the United States is entered into the United Network for Organ Sharing (UNOS) database. This database collects information about the transplant recipient and donor such as demographic information, serology testing, and human leukocyte antigen (HLA) antibody screening data. Instead of requiring study coordinators to enter this data into the EDC system a data transfer was setup with UNOS to provide the sites a dataset containing the information required for the CTOT-20 study. A separate data transfer process was setup to receive pulmonary function tests (PFTs) which are routinely collected after lung transplantation to measure lung function. Instead of study coordinators entering these tests into the EDC system data are extracted from the local PFT lab or the EHR system and sent to the statistical and data coordinating center.

Description: This talk will briefly outline the process sites use to prepare and transfer data to the statistical and data coordinating center and then will focus on how the statistical and data coordinating center cleans and monitors the data. We will also discuss the impact of using existing data sources on the quality of the data. Challenges such as mapping data to subject ID, receiving different data structures, how to transfer data securely, what level of monitoring is necessary, and how to query data that isn't received through the EDC system will be addressed. While this process is not as efficient as we had hoped and requires several manual steps to get the data from the site to the statistician for analysis it has still been more efficient than having study coordinators hand enter the data into the EDC system. We are 2 years into a 5 year study and have collected over 70,000 variables through transfers. This has freed hours of study coordinator time to focus on other study activities.

Audience: The process of transferring data to a coordinating center instead of entering it into EDC affects study coordinators, project managers, data managers, clinical monitors, and the statistician. All functional areas would find something impactful in understanding alternative avenues to collecting data and the challenges along with solutions to dealing with them. Individuals will find something to take away to help foster alternative methods of data collection in their own work environments.

This research was supported by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under Award Numbers UM2AI117870 and 5U01AI113315.

Contributors

Jamie Todd
Duke University

Laurie Snyder
Duke University

Courtney Frankel
Duke University

Jerry Kirchner
Duke University

Scott Palmer
Duke University

A66

UNDERSTANDING THE CLINICAL TRIAL ENTERPRISE'S NEWEST TREND: INCORPORATING REAL-WORLD EVIDENCE INTO RANDOMIZED CLINICAL TRIALS

LESLEY CURTIS

DUKE SCHOOL OF MEDICINE

Objectives: Describe how RWD sources can be used to support the planning and execution of randomized controlled trials

Identify barriers and potential solutions to adoption of RWE generation

Describe practical models and operational guidance for the use of RWD in randomized clinical trials to generate RWE in specific clinical trial operations activities

Introduction: RWD sources and RWE generation hold great promise for increasing the quality and efficiency of clinical trials. Panelists will discuss evidence findings to date from CTTI's work on incorporating RWE into RCTs for regulatory decision making.

Abstract Details: Real world data (RWD) can potentially support the evaluation of new medical products, new indications, and labeling changes at significantly lower cost and participant burden than with traditional randomized clinical trials (RCTs). Additionally, recent inclusion of real world evidence (RWE) provisions in the 21st Century Cures Act and Prescription Drug User Fee Act VI Commitment Letter have focused attention on using RWD sources.

Despite growing interest in leveraging RWD and RWE for clinical research, there are relatively few examples to inform such efforts—and a lack of consensus among stakeholders about appropriate methods for using RWD and RWE to support clinical trials has slowed progress in this field.

The Clinical Trials Transformation Initiative (CTTI) is addressing this gap by developing evidence-based recommendations that foster greater awareness and appropriate use of RWD sources that can be integrated into RCTs to produce RWE, specifically for the purposes of regulatory decision making.

Ultimately, the broad adoption of RWE for regulatory trials will benefit patients and other stakeholders by creating high-quality, reliable evidence to support the evaluation of medical products, new indications, and labeling changes at lower cost and participant burden than is possible with RCTs. The use of RWD also offers opportunities to answer important questions about the benefits and potential risks of medical products that are difficult or impossible to answer following traditional methodologies.

Attendees of this session will learn:

- How RWD sources can be used to support the planning and execution of randomized controlled trials.
- The barriers and potential solutions to the adoption of RWE generation.

- The concerns with RWD and RWE, how they can be addressed, and when using RWD/RWE is impractical or unwise.

- The practical models and operational guidance for the use of RWD in randomized clinical trials to generate RWE in specific clinical trial operations activities.

Contributors

Scott Evans
Harvard University

Jane Perlmutter
Gemini Group

Jack Sheehan
Janssen

Juli Tomaino
U.S. FDA

Gerrit Hamre
Clinical Trials Transformation Initiative

A67

EVIDENCED BASED REFINEMENT OF TREATMENTS – BUILDING MECHANISM EVALUATION INTO CLINICAL TRIALS

KIMBERLEY GOLDSMITH

KING'S COLLEGE LONDON

Background

Randomised clinical trials aim to answer the question "does the treatment work?", but given they have large resource requirements they should also test theoretical treatment mechanism models to address the question "how does the treatment work?" This can be done using mediation analysis to guide the refinement and development of treatments, leading to better patient outcomes. Mediation models quantify the relationship between treatment and mediator (a path) and mediator and outcome (b path), with significant a and b paths equating to significant mediation of treatment effects (Figure 1). A useful intervention evaluation paradigm refers to these as the action (a path) and conceptual (b path) theories. In a trial with no treatment effect, studying mediation can provide key treatment refinement information by addressing whether: the treatment affected its intended target mediator (the action theory was correct), and/or the mediator that was being targeted was in fact related to the outcome (the conceptual theory was correct).

We applied mediation analysis to the saMS trial of cognitive behavioural therapy (CBT) versus active supported listening for adjustment to multiple sclerosis (MS). The primary analysis showed a modest effect of CBT on distress, but not on functional impairment.

Methods

The trial outcomes were distress (General Health Questionnaire-12) and functional impairment (Work and Social Adjustment Scale). Treatment targets (potential mediators) included catastrophizing, embarrassment avoidance, Acceptance of Chronic Health Conditions scale and the Psychological Vulnerability Scale. To respect the assumption that mediation is a causal time-ordered process, the mediator measures were those taken immediately post-treatment, with outcomes measured at the primary follow-up time point (12 months post-randomisation). Mediation models were fitted using the structural equation modelling framework with 95% confidence intervals for mediated effects estimated using 1000 repetitions of percentile bootstrap.

Results

There were no statistically significant postulated mediators of the effect of CBT at follow-up, however, CBT led to significant decreases in the a paths for psychological vulnerability (standardised estimates -0.35, 95% CI -0.08 to -0.59) and catastrophizing (-0.37, -0.02 to -0.70). Increased acceptance and decreased embarrassment avoidance were significantly associated with lower distress and functional impairment (b paths, for example for functional impairment, acceptance -0.36, -0.12 to -0.63, embarrassment 0.29, 0.07 to 0.49). There were no significant mediated effects as there were no variables where both a and b paths were significant, potentially explaining modest/no effects of CBT on outcomes. These results imply that if treatment could have

greater effects on the variables with significant b paths, there may be greater effects on the outcomes.

Conclusions

We found evidence for action theory decreases in psychological vulnerability and catastrophizing after CBT treatment, and for conceptual relationships between changes in acceptance/embarrassment avoidance mediators and distress/functional impairment outcomes in MS. This suggests that refining CBT to focus more on changing acceptance and embarrassment avoidance could lead to greater effects on outcomes, improving the lives of MS patients. Such important information about how to develop and enhance treatments can be gained by including mechanistic questions and relevant measures at the trial design stage.

Contributors

Trudie Chalder
King's College London

Rona Moss-Morris
King's College London

A68

WHAT IS THE EFFECT OF TREATMENT CROSSOVER ON AN INTENTION TO TREAT ANALYSIS IN A NON-INFERIORITY TRIAL?

ISOKEN ODIA

JAEB CENTER FOR HEALTH RESEARCH

Background: In clinical trials, there are circumstances in which crossovers can occur at the beginning of treatment initiation. This may occur especially in studies where the standard care is a surgical procedure. In such circumstances, the preferred intention-to-treat (ITT) analysis becomes even more conservative when showing superiority of the experimental treatment. However, for non-inferiority studies, this approach is anti-conservative and increases the likelihood of concluding non-inferiority. Several studies have examined the effect of non-compliance and dropouts on the ITT and per-protocol (PP) analyses in non-inferiority trials, but little is known about the effect of crossovers prior to the initiation of any study treatment.

Purpose: Investigate the effect of random crossovers prior to the initiation of study treatment when analyzing non-inferiority trials with the ITT, as-treated (AS) and PP approach.

Methods: We simulated a population of 10,000 potential trial participants under the null hypothesis in which the true treatment difference was equal to the non-inferiority margin. Each participant had a continuous outcome simulated for each treatment group. A total sample size of 210 was calculated for a non-inferiority trial with effect size=0.42, one-sided alpha of 0.025, and 85% power. Samples were randomly selected (5,000 times) from the population and participants were randomly assigned to a treatment group and the corresponding simulated outcome. In the experimental treatment arm, 5%, 10% and 20% of participants were randomly selected to crossover to the standard treatment arm and their outcome was switched to the simulated standard treatment outcome. For each analysis approach (ITT, AS, or PP), computation of the 100*(1-2?) % confidence interval was calculated. The type I error was calculated as the proportion of trials in which the lower bound of the confidence interval was greater than the non-inferiority margin. All simulations were performed using SAS 9.4.

Results: For the trials in which there was 5% crossover from the experimental arm to the standard arm, the type I error rate was 3.6%, 2.6% and 2.5% for the ITT, AS and PP analyses, respectively. For 10% crossover, the rate was 4.9%, 2.5% and 2.4%. For 20% crossover, the rate for the ITT increased to 8.7% while the rate remained similar (2.7% and 2.5%) for the other approaches (AS and PP).

Conclusions: Selecting an analysis population in non-inferiority trials is an issue that is still debated. This preliminary analysis suggests that the ITT approach had the highest type I error rate in all scenarios of random crossover. We will quantify the bias under the different scenarios and conduct further analysis to explore non-random crossover, varying effect sizes, and varying power.

Contributors

Michele Melia
JAEB CENTER FOR HEALTH RESEARCH

A69

ETHICAL ASPECTS OF CLINICAL TRIALS IN INDIA: A SYSTEMATIC REVIEW OF RESEARCH IN THE PAST TWO DECADES

SANGEETHA PARAMASIVAN

UNIVERSITY OF BRISTOL

Background

International clinical trials (CTs) that recruit participants from low- and middle-income countries (LMICs) are rapidly becoming the norm for economic, pragmatic and scientific reasons. However, unacceptable ethical practices, such as failure to obtain participants' informed consent, have been frequently reported in LMICs over the past few years. In India, CTs were briefly halted in 2013 in response to concerns for patient autonomy and safety. With the anticipated upsurge in the global proportion of CTs conducted in India, we systematically reviewed the evidence to a) obtain an overview of the research conducted in relation to ethical aspects of conducting CTs in India and b) investigate key themes from this evidence base, to inform future research.

Methods

A comprehensive search strategy combining terms related to ethics/bioethics, informed consent, CTs and India was developed and applied by two authors across nine databases (Medline, Embase, PsycInfo, IBSS, Scopus, CINAHL, Cochrane, Web of Science and ORRCA), up to September 2017.

Primary/secondary research studies in India that explored ethical aspects of CTs, with any key stakeholder groups (patients/public, professionals) were included. Animal studies, commentaries and studies outside India were excluded, with no restrictions based on study design, language or age.

Title/abstract screening was carried out by one author, with 20% of articles blind screened by another to check for accuracy. Full text screening and data extraction of included studies were carried out by two authors, with discrepancies discussed and resolved. Quantitative and qualitative studies were analysed separately and drawn together to provide a detailed picture of the available evidence.

Results

A total of 10840 articles were identified, titles/abstracts of 6508 articles were screened after removing duplicates, full text was obtained for 255 articles and 138 were included in the review.

Preliminary findings suggest that research on ethical aspects of CTs in India has mainly occurred over the past two decades and spans a wide range of topics, population groups and study designs. The most researched area was knowledge, attitudes and practices (KAP) of patients/public and professionals (healthcare staff/students, researchers) on topics such as informed consent for CTs. Other key areas studied include functioning of research ethics committees and reporting of ethical

aspects of CTs in publications and trial registries. Studies investigating what transpires in a CT informed consent appointment were particularly lacking.

The KAP studies, with the exception of a few, consistently reported several misconceptions in relation to CTs amongst lay and professional groups. Of particular interest is that doctors, healthcare trainees (medical/nursing students) and medical/dental faculty members were reported to have poor knowledge of the ethical conduct of CTs.

Conclusion

This review demonstrates that while a wide range of ethical areas in CTs have been studied in India, the focus appears to be on assessing knowledge levels. There is a need to move towards examining the components of the informed consent procedure, such as the doctor-patient interaction in a consent appointment. The inadequate knowledge levels among professional groups emphasises the need for mandatory ethics training for clinicians and researchers at pre- and post-qualification levels.

Contributors

Philippa Davies
University of Bristol

Alison Richards
University of Bristol

Nicola Mills
University of Bristol

Leila Rooshenas
University of Bristol

Julia Wade
University of Bristol

A70

ADAPTIVELY MONITORING CLINICAL TRIALS WITH SECOND-GENERATION P-VALUES

JONATHAN CHIPMAN

VANDERBILT UNIVERSTIY

The Food and Drug Administration is committed to "facilitate the advancement and use of complex adaptive, Bayesian, and other novel clinical trial designs" - Prescription Drug User Fee Act VI. We introduce a novel design based on the Second-generation p-value (SGPV). Recently introduced at the ASA Symposium on Statistical Inference, the SGPV indicates when the data are compatible with the alternative hypothesis, the null hypothesis, or when inconclusive. This requires that the minimum clinically meaningful effect size is specified upfront. Our design uses this information to reduce the false discovery rate by ignoring statistically significant results for clinically meaningless effect sizes. SGPVs outperform traditional approaches based on adjusting the p-value for multiple comparisons or looks. Our novel design permits sequential monitoring and interim examinations of multiple endpoints. The trial halts when the data support either convincingly superior or uninteresting clinical results. In extensive simulations and the currently active clinical trial REACH, we compare our method's performance in terms of error rates, false discovery rates, bias, and average stopping times to traditional adaptive designs such as the SPRT, Conditional SPRT, and group sequential designs, Bayesian Adaptive and Simons Two Stage Designs, and OBrien-Fleming Boundary Stopping Rules.

Contributors

Robert Greevy, Jr.
Vanderbilt University

Lindsay Mayberry
Vanderbilt University

Jeffrey Blume
Vanderbilt University

A71

USING FDA WEBSITE AND CLINICALTRIALS.GOV TO IDENTIFY TRIALS FOR NETWORK META-ANALYSIS:

A CASE STUDY ON FIRST-LINE MEDICATIONS FOR GLAUCOMA

LIN WANG

JOHNS HOPKINS BLOOMBERG SCHOOL OF PUBLIC HEALTH

Using FDA website and ClinicalTrials.gov to identify trials for network meta-analysis:

A case study on first-line medications for glaucoma

Authors: Lin Wang, Yong Chen, Christopher Schmid, Tianjing Li

Background: Network meta-analysis (NMA) extends pairwise meta-analysis by synthesizing both direct evidence within trials and indirect evidence across trials. NMA can address a broader research question than pairwise meta-analysis by comparing all interventions for a given condition in a single analysis. However, identifying, appraising, and analyzing all relevant data is resource-intensive and time-consuming. Bibliographic databases such as PubMed, Embase and Cochrane CENTRAL are almost always searched to identify published trial reports. For regulated products (e.g., pharmaceuticals and biologics), the approval packages available from the US Food and Drug Administration (FDA) website contain valuable information about trials, although not all of them are published. ClinicalTrials.gov is another data source that can be tapped into in identifying trials and their summary results.

Objective: To compare the use of bibliographic databases, the FDA website and ClinialTrials.gov to identify trials for NMAs of drug interventions in a specific case study; and to assess how inferences might be affected by using different combinations of these information source.

Method: Building upon a recent NMA and available data set, we searched the FDA website and ClinicalTrials.gov for randomized controlled trials on first-line medications for open angle glaucoma. Two individuals independently screened, selected studies, and extracted data from relevant records. When a trial was identified in multiple sources, we compared information presented from the different data sources. We fitted random effects NMA models and analyzed five networks of trials for the outcome of intraocular pressure (IOP) at 3 months: all unique trials; published trials only (existing data); FDA trials; ClinicalTrials.gov trials; published trials but not identified from the FDA website and ClinicalTrials.gov. We contrasted findings from these analyses.

Results: We identified 114 published trials, 31 FDA trials (14 published and 17 unpublished), and 27 ClinicalTrials.gov trials (16 published and 11 unpublished). Published trial reports provided most information about trial design, statistical methods, and results, while ClinicalTrials.gov records provided the least. Information about trial characteristics and risk of bias sometimes differed across sources, and conflicting information was difficult to disentangle. 107/114 published trials (17340/21991 participants), 16/31 FDA trials (5454/10145 participants), and 9/27

ClinicalTrials.gov trials (2299/11861) provided sufficient data for NMA (a point estimate and its precision). The FDA trials were larger and more homogeneous clinically, methodologically, and statistically compared to trials identified elsewhere. The effect estimates generally agreed when different sources of data were used for NMA, although the precisions varied (Figure).

Conclusions: The FDA website is useful for identifying trials for NMA for regulated products. In our case example, NMA based on FDA trials alone provided reasonably precise estimates of relative effects. Reporting of trial design and results can be improved in both the drug approval packages and on ClinicalTrials.gov.

Figure. Treatment effect estimates based on different networks of trials

Contributors

Yong Chen
University of Pennsylvania Perelman School of Medicine

Christopher Schmid
Brown School of Public Health

Tianjing Li
Johns Hopkins Bloomberg School of Public Health

A72

DIGITAL STRATEGIES FOR MAKING CLINICAL TRIALS MORE ACCESSIBLE: SWOG'S WEBSITE REDESIGN

WENDY LAWTON

SWOG GROUP CHAIR'S OFFICE, OREGON HEALTH & SCIENCE UNIVERSITY

Cancer clinical trials offer patients unique treatment options, including targeted therapies and immunotherapies. Yet research shows that only 3 to 5 percent of American adults with cancer enroll in a clinical study.

As a result, entities including the American Society of Clinical Oncologists, the National Cancer Institute, and the American Cancer Society have launched cancer clinical trial education campaigns in recent years. At the same time, the world's largest clinical trials search tool, ClinicalTrials.gov, is undergoing an extensive overhaul to make it easier for the public to use.

SWOG also has a goal of making its trials more accessible, and understandable, to the public it serves. A global cancer clinical trials network funded by the National Cancer Institute, SWOG in 2017 redesigned and rebuilt its extensive website, SWOG.org, along with the web development firm iFactory.

New and improved features were designed to make trials easier to find and understand:

* Intuitive trial search tool that allows trials to be searched based on location, disease, etc.

* Intuitive publications search tool on trial results presented in journals and at scientific meetings

* "Patient resources" section with an explanatory video and links to patient-centered resources

* Mobile responsive design that allows for ease of use on tablets and smartphones

* Inviting, human-centered design and plain language text

* Patient photos and testimonials from people who've joined SWOG trials

* Glossary of cancer and cancer research terms

* "How We Work" section that explains the tasks and roles involved in creating and running trials

In addition, SWOG and iFactory built in several site improvements – including a new content management system and improved analytics – that will allow for fast and effective changes to the site based on usage and member requests.

Since the site launched Dec. 1, 2017, member reaction has been positive, particularly the improved search tools, patient stories, and "How We Work" section. SWOG will continue to solicit member feedback, and use that to make annual improvements and updates in order to keep the site relevant, understandable, and easy to use.

Acknowledgement: The project reported in this publication was supported by a grant from The Hope Foundation (philanthropic arm of the SWOG Cancer Research Group), and by the National Cancer Institute of the National Institutes of Health under Award Numbers CA180888 and CA180819. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Contributors

Chris Cook
SWOG Statistics and Data Management Center

Morgan Cox
The Hope Foundation

Courtney Wille
SWOG Group Chair's Office, Oregon Health & Science University

Keith Goodman
SWOG Statistics and Data Management Center

Rick Dai
SWOG Statistics and Data Management Center

A73

A "COMPOUND" PROBLEM: GENERAL STRATEGIES AND KEY RESOURCES FOR THE UTILIZATION OF APPROVED, COMMERCIALLY AVAILABLE DRUGS IN A BLINDED, CONTROLLED CLINICAL TRIAL

ERIC HARDTER

THE EMMES CORPORATION

Adequate and well-controlled clinical trials (AWC) are the gold standard for the determination of the efficacy and safety of an intervention (drug or biologic), with blinded, placebo or active control studies at the forefront. To establish that a comparison across treatment groups is valid and that a blind is maintained, it is essential to ensure the identity, potency, and purity of any investigational products (IP) and that adequate, compliant manufacturing processes are in place to maintain the blind at the IP level. Several key considerations need to be accounted for during the protocol development and planning stages to safeguard against common logistical and regulatory pitfalls of using approved, commercially available medication in a controlled clinical study.

IP characteristics, including physical and chemical specifications and route of administration, will drive downstream decision making regarding compounding to enable blinded medication to be produced, including the extent of manipulation, impact on sterility, and the production of a matched placebo. Projected recruitment totals, enrollment windows, and randomization schemes will inform the planning of IP and placebo sourcing/production, and study product stability testing schedules. Identification of a suitable manufacturer will assure that clinical trial needs can be met and product disposition can be facilitated, and subsequent testing provides confidence that the IP and placebo/active comparator are prepared in conformance with current Good Manufacturing Practices (cGMPs). Finally, for studies regulated by the Food and Drug Administration (FDA) under an Investigational New Drug (IND) application, intimate knowledge of preliminary and continuing reporting obligations will enhance the likelihood of ongoing Agency approval and shield against the prospect of a clinical hold.

The Emmes Corporation has supported several blinded, placebo- and active comparator-controlled clinical trials, including those operating under IND regulations. These studies have spanned numerous disease areas and proposed indications, and have included IPs ranging from small molecule drugs to biological vaccines. In providing assistance for such a diverse array of trials, Emmes has developed best practices for ensuring compliant IP and placebo manufacture suitable for the stage of development and the goals of the study. Additionally, this scope of understanding has included how best to troubleshoot the various hazards encountered, both anticipated and unforeseen, from prior to study implementation through site closeout.

This presentation will highlight a logical and temporal series of considerations for study product manufacturing in blinded, controlled trials, including but not limited to those described above, as well as emphasize the strategies and resources necessary to meet and exceed the needs of these complicated and pivotal studies.

Contributors

Gillian Armstrong
The Emmes Corporation

Matthew Wright
The Emmes Corporation

Dikla Shmueli-Blumberg
The Emmes Corporation

Eve Jelstrom
The Emmes Corporation

Robert Lindblad
The Emmes Corporation

A74

SAFETY ASSESSMENT COMMITTEE JOINS WITH DATA MONITORING COMMITTEE - SAFETY IS THE NEW EFFICACY

WEI WANG

AXIO RESEARCH, LLC

In the clinical trial world, it is well-known and well-established practice for many sponsors to have a Data Monitoring Committee (DMC), a group of independent clinicians and biostatisticians, to monitor the ongoing clinical trial(s) to protect patients' safety and preserve the validity and scientific merit of the trial conducts.

In late 2015, FDA released a new guidance on safety assessment committees (SAC) for IND safety reporting. SAC is also charged with monitoring the safety of the clinical trial participants, but SAC is from a different perspective and different guidelines comparing to DMC.

Despite it is still a very new guidance, and many sponsors are working to find out a way of how to properly implement SAC, the author's organization has had the first-hand experience on organizing and supporting SAC for some top pharmaceutical companies. What is the author's experience with SAC support? What are the special statistical methods used in SAC report? What are the differences between setting up a DMC and a SAC? What are the challenges in SAC organization? These topics will be discussed.

Contributors

David Kerr
Axio Research, LLC

A75

USING COMPOSITE SURROGATE BASELINE VARIABLES IN A MODIFIED COVARIATE-CONSTRAINED RANDOMIZATION ALGORITHM

KATHRYN JACKSON

NORTHWESTERN UNIVERSITY

Background: Successful control of covariate imbalance through randomization requires a manageable number of variables and accurate variable measurement. The nature of large-scale practice facilitation projects can make accurate and timely collection of baseline data challenging, and controlling imbalance across arms may involve several baseline covariates. In these cases, use of a "surrogate" baseline variable in randomization may be useful.

The Healthy Hearts in the Heartland (H3) project aims to understand how practice facilitation strategies can promote cardiac health through improvement of four cardiac measures at the practice level. Here we evaluate the use a derived score based on estimates of these four variables as a proxy for the true values of these variables at baseline in a modified constrained randomization algorithm.

Methods: Ideally randomization would involve the true values of each of the four variables, but logistical constraints of the study required estimation for each at baseline. To efficiently control imbalance in all four variables, we used a composite surrogate measure in randomization:

1) Initial estimated values for each of the four measures were obtained

2) Empirical distributions for each measure were divided into tertiles and scored 0, 1, or 2

3) The average of these scores served as the composite score for each practice included in the randomization algorithm

We employed a modified constrained randomization scheme to control imbalance in this surrogate composite score across two arms (Nietert, 2009; Zhao, 2012). Post randomization, we captured more robust (validated) individual-level measurements for baseline and calculated a composite score for each practice as outlined above. Success of the algorithm utilizing the surrogate composite score was ultimately measured via the comparison of these final, validated, individual measurements across arms. The Wilcoxon Rank Sum (WRS) test compared median values for each measurement across arm, and the Wilcoxon Signed-Rank (WSR) test compared initial surrogate composite scores to those based on the validated measurements. Sample correlation coefficients assessed association of the overall surrogate composite from randomization with the four validated measurements.

Results: Randomization occurred in 226 practices across four "waves" with equal allocation across arms. The randomization algorithm adequately controlled imbalance in the composite surrogate score (p=0.661) and in the composite based on the validated measurements (p=0.691). This resulted in adequate balance for each of the four true outcomes at baseline: median value 0.78 vs. 0.77 (p=0.428); 0.64 vs. 0.64 (p=0.634); 0.78 vs. 0.75 (p=0.187); 0.83 vs. 0.85 (p=0.336). We found

no significant differences between the composite score based on the initial estimates and that based on validated estimates (WSR p=0.708). The surrogate composite score was significantly correlated with the post-randomization score (Spearman's r=0.392, p<0.001) and with all except one validated variable (r=0.223, 0.315, 0.053, 0.245; p=0.001, <0.001, 0.451, <0.001, respectively).

Conclusions: Using this modification to constrained randomization, we utilized a surrogate composite score based on estimated baseline outcome values to achieve adequate baseline balance across arms for H3's main study outcomes. This approach may be adapted and implemented for similar studies in which individual variables are not readily available at the time of randomization.

Contributors

Theresa Walunas
Northwestern University

David Liss
Northwestern University

Stephen Persell
Northwestern University

Jody Ciolino
Northwestern University

A76

INTRODUCING A NEW STATA PACKAGE FOR SAMPLE SIZE CALCULATION IN TRIALS WHERE THE OUTCOME IS A RATE

STEPHEN NASH

LONDON SCHOOL OF HYGIENE AND TROPICAL MEDICINE

Motivation

Sample size is a critical design consideration when planning a randomized controlled trial (RCT). Where the aim of an intervention is to slow disease progression the natural outcome is a rate of change in some continuous variable. Rates of change can be estimated more precisely when outcomes are measured at interim time-points as well as at randomization and at final follow-up and can be compared between the treatment groups in a linear mixed model analysis. It is not straightforward to estimate sample sizes for such trials because the variance of the treatment effect depends on the proposed length of the trial and the frequency and timing of interim visits as well as the sample size, and the parameters defining this dependency are usually unknown. In such situations it is possible to use existing data from longitudinal studies to estimate these parameters and hence inform sample size.

Purpose

We introduce a new Stata package that translates existing methodology(1,2) in this area into a user-friendly program for sample size calculation, appropriate for planning 2-arm parallel trials of treatments that are expected to slow disease progression by a constant amount throughout follow-up.

Methods and Features

Our package fits a linear mixed model to a user-supplied longitudinal dataset, extracting relevant parameters to estimate mean rates of changes and components of between- and within-person variability. These estimates, together with other user inputs, are then used to calculate the required sample size for a proposed RCT. Users specify the visit schedule, expected loss to follow up rates, Type I and Type II error rates, and the effectiveness of the treatment.

Either observational or trial data are accepted for model fitting. Observational data must include subjects with the disease under investigation, and can optionally include a group of healthy people without disease. In the latter case the components of variability used in the sample size calculation are estimated only from subjects with disease. Trial data should include subjects from both treatment and control arms; here components of variability are estimated from both trial arms and it is assumed that the mean rates of change in the control arm of the proposed trial will be the same as in the control arm of the existing trial.

The program allows three scenarios regarding treatment effectiveness: 1) it is expected to reduce the rate of change by a certain proportion, appropriate when a 100% effective treatment would halt change altogether (suitable for observational data without healthy people, or for trial data); 2) it is

expected to reduce the excess rate of change (the rate of change over-and-above that seen in a disease-free population) by a certain proportion, appropriate when a change over time is expected to remain even with a 100% effective treatment (suitable for observational data that includes healthy people without disease); or 3) it is expected to reduce the rate of change by the same amount as observed in a previous trial (suitable for trial data from an RCT of the same, or a similar, intervention).

Contributors

Katy Morgan
London School of Hygiene and Tropical Medicine

Chris Frost
London School of Hygiene and Tropical Medicine

Amy Mulick
London School of Hygiene and Tropical Medicine

A77

UTILIZING A ROBUST AND COMPREHENSIVE LOCATOR FORM AS A RETENTION TOOL IN CLINICAL TRIALS

DAGMAR SALAZAR

THE EMMES CORPORATION

Background:

The National Institute on Drug Abuse (NIDA) Clinical Coordinating Center (CCC) and Data and Statistics Center (DSC) support the National Drug Abuse Treatment Clinical Trials Network (NDAT CTN), which evaluates behavioral, pharmacological, and integrated treatment interventions for substance use disorders. These clinical trials routinely enroll participants who may be particularly difficult to find during treatment and follow-up. One tool for increasing retention has been the use of locator forms to obtain contact information from study participants to facilitate locating them. Standard forms collect information such as the participant's current address, email address, and phone numbers. However, in NDAT CTN studies, which frequently involve participants with unstable living conditions conventional locator forms were not sufficient to reach participants.

Objective:

The goal was to create a comprehensive and robust locator form to more successfully locate participants for their study visits, and to make this tool a focal assessment at screening/baseline.

Methods:

The CCC, DSC, and the study investigative teams have refined the content and collection method of the locator form through numerous iterations, across clinical trials. Some best practices are reviewed, which may aid in locating particularly hard to find participants.

Results/Conclusion:

When a participant falls out of contact with the study team, it is important to use all information available to regain contact and avoid a missed study visit. Use of less traditional information and methodology may be useful. For example, participant "Facebook" profile information (or other social media outlets, such as Twitter) should be collected in the locator form. Similarly, the form could include areas participants would frequent (e.g., medical, social services, support groups, churches, bars) or sleep if unstably housed (e.g., shelters, parks) to aid attempts to locate hard to find participants in the field. Collection of social security and driver's license numbers are useful and can aid in searches of public records. Information shared during the completion of other study assessments may also inform the collection of locator information. The more robust and inclusive a locator form is, the greater the chance to find a participant and retain them in a study. The collection of locator information should be viewed as an ongoing dialogue with the participant, with a goal of not only obtaining as much information as possible, but obtaining the essential information needed to find that specific participant. Reviewing and updating locator form information at each participant interaction is also important to maintain complete and accurate

information. Furthermore, leveraging the data system to enforce collection of information can increase the usefulness of the tool. For CTN studies, flags and requirements were programmed into the electronic data capture system to ensure study team collection of all required information. To ensure confidentiality, data collected on the form was encrypted so that it was only visible to the study team at the site and not transferred into the data files.

Contributors

Phoebe Gauthier
Geisel School of Medicine at Dartmouth

Mitra Lewis
The Emmes Corporation

Patricia Novo
New York University School of Medicine

Lauren Yesko
The Emmes Corporation

Jeremy Wolff
The Emmes Corporation

A78

USE OF A BAYESIAN APPROACH TO INCORPORATE TOXICITY CONSIDERATION INTO THE TWO STAGE DESIGN OF PHASE II CLINICAL TRIALS

CHIN-FU HSIAO

NATIONAL HEALTH RESEARCH INSTITUTES

Introduction/Objective. The primary objective for the phase II trial is to be concerned with safety and efficacy estimation. Therefore, The objective of this study is to develop two Bayesian two-stage designs for Phase II clinical trials with incorporation of toxicity consideration.

Background. The majority of current cancer phase II clinical trialsapproaches are frequentist statistical approaches and only concerned with efficacy estimation.

Methods. A single arm Phase II clinical trial will be conducted to simultaneously evaluate the safety and efficacy of a new drug. There will be 4 parameters: probabilities of observing clinical responses without experiencing toxicity, observing clinical responses with experiencing toxicity, observing no clinical response without experiencing toxicity, and observing no clinical response with experiencing toxicity. We use a Dirichlet distribution as a prior.

Results.Two designs are proposed. The first design is based on the probability of observing patients who will have clinical response and will not experience toxicity, while the second design is based on both the probability of observing patients who will have clinical response and the probability of observing patients who will not experience toxicity.

Conclusions. Two Bayesian two-stage designs for Phase II clinical trials with incorporation of toxicity consideration were developed. Sample size determination for each stage is evaluated.

Contributors

Yu-Chieh Cheng
National Health Research Institutes

Hsiuying Wang
Nation Chiao Tung University

Toshimitsu Hamasaki
National Cerebral and Cardiovascular Center

A79

ETHICAL IMPLICATIONS FOR PRAGMATIC TRIALS INVOLVING HEAD-TO-HEAD COMPARISONS OF USUAL CARE INTERVENTIONS: AN ETHICAL ANALYSIS OF THE MINT TRIAL

CORY GOLDSTEIN

ROTMAN INSTITUTE OF PHILOSOPHY, WESTERN UNIVERSITY

The Myocardial Ischemia and Transfusion (MINT) trial is a large, National Institutes of Health funded randomized controlled trial (RCT) designed to determine whether one of two red blood cell transfusion strategies commonly used in clinical practice reduces death or nonfatal heart attacks in patients who have suffered a heart attack and are anemic. This trial will ultimately randomize 3,500 hospitalized patients across the United States and Canada to receive either a liberal or a restrictive transfusion strategy.

In August 2017, three months after enrollment in the MINT trial began, Public Citizen filed a complaint to the Office for Human Research Protections calling for the immediate suspension of patient enrollment. The ethical issues identified by Public Citizen were (1) the protocol lacked an adequate description of the current state of evidence and practice, and the risks of a restrictive transfusion policy; and (2) the consent document lacked an adequate description of the research purpose, current state of evidence and practice, and reasonably foreseeable risks of study participation.

We conducted an ethical analysis of the MINT trial. In our analysis, we argue that the central ethical question raised by the MINT trial is: are usual care interventions in pragmatic RCTs properly conceived of as practice or research? In this talk, we consider arguments proffered for conceptualizing usual care interventions as practice, but ultimately argue that usual care interventions commonly adopted in pragmatic RCTs ought to be conceptualized as research. This is because randomization and protocolization ultimately undermine the physician's duty of personal care. When patients are deprived of the protection of the individualized judgment of their physician, review by research ethics committees is needed to protect the liberty and welfare interests of patients participating in research.

Our ethical analysis has various implications for pragmatic trials involving head-to-head comparisons of usual care interventions. (1) Usual care interventions are rightly classified as part of the research and, as such, they fall under the purview of the research ethic committee. (2) Usual care interventions must be carefully considered as part of the research ethics committee's benefit-harm analysis. A systematic review of the evidence should support the conclusion that clinical equipoise exists. (3) Usual care interventions, including their benefits, risks and alternatives, must be disclosed in the research informed consent process, unless conditions for an alteration or waiver of consent obtain. Simplified (or targeted) approaches to informed consent are desirable as they are consistent with both respect for participant autonomy and efficient RCT recruitment.

Contributors

Charles Weijer
Rotman Institute of Philosophy, Western University

A80

DECOMPOSITION OF THE TREATMENT EFFECT ESTIMATOR IN STEPPED WEDGE TRIALS: UNDERSTANDING THE HORIZONTAL AND VERTICAL CONTRIBUTIONS

ANDREW FORBES

MONASH UNIVERSITY

Introduction

A linear mixed model incorporating a random cluster effect is the most commonly used model for analysis of complete stepped wedge designs with Gaussian outcomes and a repeated cross-sectional sampling structure. It is recognised that the maximum likelihood estimator of the treatment effect in this model is a combination of horizontal (within cluster) and vertical (between cluster) comparisons. However, the precise nature of this combination has not previously been clearly articulated for these designs.

Methods

We apply standard results using partitioned matrices to derive a simple expression for the weighted combination of the horizontal and vertical components of the treatment effect estimator, each presented as linear combinations of cluster-period means. We extend the mixed model to incorporate random effects appropriate for a closed cohort design and derive the analogous results under this design.

Results

The weights assigned to the horizontal and vertical comparisons involve a simple expression depending on the number of periods in the design, the cluster size and the intra-cluster correlation. We use this result to describe scenarios in which the treatment effect estimator is dominated heavily by the horizontal comparisons. We provide explicit expressions for the horizontal and vertical components of the treatment effect estimator in a number of example designs and explain the intuition behind them. We also describe how the decomposition provides a basis for the construction of randomisation tests. The extension to the closed cohort design involves identical horizontal and vertical components as the cross-sectional sampling design, the only difference being in the construction of the weights.

Conclusions

The decomposition into horizontal and vertical components enables a better understanding of the explicit linear combinations of cluster-period means underlying the treatment effect estimator. It also describes where the maximal information resides in these designs, leading to suggestions for optimal incomplete designs.

Contributors

JNS Matthews
Newcastle University

A81

IMPLEMENTING STANDARDIZED ASSESSMENTS IN EDC: BENEFITS AND CHALLENGES BASED ON EXPERIENCE OF INTEGRATING THE PHENX TOOLKIT INTO A CLINICAL TRIALS NETWORK

JEREMY WOLFF

THE EMMES CORPORATION

Introduction

The PhenX (consensus Measures for Phenotypes and eXposures) Toolkit is a collection of high-quality standard measures for use in biomedical research that span a number of domains (www.phenxtoolkit.org). One subset of the Toolkit is the Core Tier 1 measures of the Substance Abuse and Addiction (SAA) Collection, which include 19 measures related to the use of alcohol, tobacco and other substances, demographics, body mass index, quality of life and HIV testing history. As a method of standardization of data across clinical trials, the National Drug Abuse Treatment Clinical Trials Network (CTN) of the National Institute on Drug Abuse (NIDA) encourages investigators to incorporate these Core Tier 1 SAA measures into their CTN studies.

Objective

The Emmes Corporation in its role as the Data and Statistics Center (DSC) of the NIDA CTN for the last 8 years is responsible for implementing efficient and standardized data management systems for CTN trials. In order to assist investigators in incorporating Tier 1 measures into their studies, the DSC needed to assess the nuances of the measures and then plan their implementation into an electronic data capture (EDC) system.

Methods

Implementation of these measures in our EDC system required frequent review of the PhenX website, thoughtful conversion of paper Case Report Forms (CRFs) into questions on an eCRF consistent with the network's existing eCRF style guide, incorporation of complex parenting and skip logic, contacting the PhenX group for clarification and review with investigators to ensure study needs were being met.

Results

PhenX Core Tier-1 measures have been incorporated in 7 CTN studies across 47 sites over the past X years. Of the 19 Core Tier-1 measures, 3 were straightforward to implement across all studies. However, the remaining 16 presented challenges during implementation that included resistance from the investigators in using the standard tools that were considered not optimal to address research questions, ensuring duplicative data were not collected based on other assessments that were integrated into the study, considering mapping from similarly worded questions from other data collection instruments, implementing complex skip logic and creating a clear paper source CRF, and keeping up with frequent changes to the PhenX measures. Specific examples of our

interpretation of the measures as specific eCRFs, challenges and methods of resolution will be presented.

Conclusions

Rapid, flexible and consistent implementation of Core Tier-1 measures across multiple NIDA CTN trials accomplishes the PhenX goal of collaborative research and consistent collection of high-quality data in a standardized manner across multiple studies. However, it requires careful planning, testing, review, problem-solving and communication with internal and external stakeholders. Lessons learned here with the Tier-1 measures can be applied to other PhenX collections, and other instruments that may need to be standardized in a clinical trials network setting. Ultimately, data collected in a standardized format allows for data to be easily understood and combined across multiple studies, leading to more impactful and generalizable conclusions, both for NIDA CTN investigators as well as researchers using these data when they become publically available through NIDA DataShare.

Contributors

Dikla Shmueli-Blumberg
The Emmes Corporation

Lauren Yesko
The Emmes Corporation

Anne Hoehn
The Emmes Corporation

Jennifer McCormack
The Emmes Corporation

Paul VanVeldhuisen
The Emmes Corporation

A82

A SYSTEMATIC ANALYSIS OF UK NATIONALLY-FUNDED SURGICAL PILOT AND FEASIBILITY STUDY PROTOCOLS FROM THE LAST 10 YEARS TO INFORM AND OPTIMISE FUTURE SURGICAL TRIALS.

KATHERINE FAIRHURST

UNIVERSITY OF BRISTOL

Background

High quality randomised controlled trials are necessary to inform evidence-based surgical practice. These may be optimised by well performed pilot and feasibility studies that explore uncertainties around trial design and conduct. However, the published literature is littered with examples of underpowered surgical trials, mislabelled as pilot studies. Detailed understanding of how and when to design and perform pilot and feasibility studies in surgery is lacking. The aim of this work is to identify key design features in successfully funded surgical pilot and feasibility studies, to inform guidance in this area.

Methods

Pilot and feasibility studies of surgical interventions funded by the National Institute for Health Research (NIHR) Health Technology Assessment (HTA) and Research for Patient Benefit (RfPB) programmes between 2005 and 2015 were identified. Both are established major UK programmes, funding studies up to £10m (HTA) and £350k (RfPB). Original study protocols and associated published outputs were sourced. Internal pilot studies, which are designed and funded as part of the main trial and should thus be considered separately (as other methodological work in this area has done), were excluded. Systematic reviews were also excluded as none met the definition of pilot/feasibility work by stating an intention to inform a future main study. Data extracted from the protocols included characteristics of the study type and design and the rationale cited for performing the work, including perceived uncertainties around conducting a definitive main trial. Descriptive statistics are presented.

Results

Over the 10-year funding period, n=1341 studies were identified. Of these, n=35/1341 (2.6%: RfPB n=25/35, 71%; HTA n=10/35, 29%) were eligible for inclusion. Of the included studies, most (n=29/35, 83%) were randomised, with over half of these (n=15/29, 52%) also including other types of pre-trial work, such as qualitative interviews, participant surveys and economic modelling. The n=6/35 (17%) non-randomised studies varied in design, including national audits, non-randomised cohort studies, systematic reviews and questionnaire surveys and/or qualitative work (e.g. interviews/focus groups) to explore stakeholders' opinions. Addressing uncertainties around recruitment was the most commonly cited rationale for conducting pre-trial work (n=30/35, 85.7%). Only one in four studies (n=9/35, 25.7%) sought to address uncertainties surrounding intervention stability, implementation or delivery. One in three studies (n=11/35, 31.4%) stated an intention to collect 'preliminary' data on safety and/or effectiveness, of which n=8/11 (73%)

specifically planned to statistically compare intervention(s) and/or control groups to test effectiveness and/or safety. Nearly half (n=14/35, 46%) of all studies had not made their results publicly available to date.

Conclusions

This analysis of peer-reviewed nationally-funded surgical pilot/feasibility studies indicates that, while methodological issues are explored to some extent, the full potential of pre-trial work to inform and optimise definitive surgical studies is likely not being realised. In addition, much pilot/feasibility work remains unreported. Future work should focus on engaging key stakeholders to develop recommendations and guidance on the design and optimal use of pilot/feasibility work to inform surgical trials.

Contributors

Jane Blazeby
Centre of Surgical Research & Medical Research Council (MRC) ConDuCT-II Hub for Trials Methodology Research (Collaboration and innovation for Difficult and Complex randomised controlled Trials In Invasive procedures) Bristol Medical School, Department of

Ceri Rowlands
Centre of Surgical Research & Medical Research Council (MRC) ConDuCT-II Hub for Trials Methodology Research (Collaboration and innovation for Difficult and Complex randomised controlled Trials In Invasive procedures)Centre of Surgical Research & Medical R

Shelley Potter
Centre of Surgical Research & Medical Research Council (MRC) ConDuCT-II Hub for Trials Methodology Research (Collaboration and innovation for Difficult and Complex randomised controlled Trials In Invasive procedures) Bristol Medical School, Department of

Carrol Gamble
Department of Biostatistics & Medical Research Council (MRC) NWHTMR (North West Hub for Trials Methodology Research), University of Liverpool, UK.

Kerry Avery
Centre of Surgical Research & Medical Research Council (MRC) ConDuCT-II Hub for Trials Methodology Research (Collaboration and innovation for Difficult and Complex randomised controlled Trials In Invasive procedures) Bristol Medical School, Department of

A83

RISK-BASED VALIDATION IN CLINICAL TRIAL REPORTING: FOCUS ON WHAT MATTERS MOST

AMBER RANDALL

AXIO RESEARCH

Quality control is fundamental to ensuring both correct results and sound interpretation of clinical trial data. Most validation procedures are a function of regulatory requirements, industry standards, and corporate philosophies. However, in the current environment of lean work forces and increased workload, the traditional approach of 100% independent programming may no longer be feasible. Further, we would argue that full double programming may not be the best approach for maintaining high quality. Instead we propose the implementation of a risk-based validation system in which the validation method is chosen individually for each piece of a project. Risk assessment is based on matrix of factors and the outcome guides the assignment of a testing methodology appropriate for that particular risk level. This ensures that higher effort and rigor is focused in areas that matter most.

In order for risk-based approaches to work, it is critical that someone with the appropriate experience be involved in the initial risk assessment. Additionally, it is critical to allow the risk level and testing methodology to be modified after initial programming is performed. Far too often programs require unanticipated modification to accommodate incomplete or unexpected source data.

In a risk-based environment it is also important to develop metrics to evaluate its performance. Key information needs to be captured throughout the programming and testing process to improve risk assessment and assignment of appropriate testing methodology.

We will discuss our proposal for the application and implementation of a risk-based validation system and explore some of the tools that make such an approach possible.

Contributors

William Coar
Axio Research

A84

RETHINKING THE ARCHITECTURE OF PRAGMATIC TRIALS

CLARE RELTON

QUEEN MARY'S UNIVERSITY

Introduction

In an ideal world, healthcare systems would continually learn from the healthcare they deliver. The randomised controlled trial (RCT) has unsurpassed advantages in generating learning that is free from confounding. But the current trials system is grossly inefficient in generating this learning. Over the last decade a number of new non-standard approaches to pragmatic trial designs have been created. This review describes these designs and how they may help healthcare systems conduct more efficient and informative trials.

Methods

To identify these new approaches we conducted a review of MEDLINE since 2007 and consulted experts in the field. For the purposes of this review we focussed on non standard pragmatic trial designs which randomise individuals. We excluded cluster RCTs and trial designs with no published examples. We describe: (i) key features of these designs, (ii) what is known about their data quality and efficiency, and (iii) ongoing endeavours in the international research community to address the questions raised by these non-standard approaches.

Findings

We identified 5 main types of trial designs: Trials within Cohorts (Relton et al, 2010), Trials within Trials using Electronic Health Registers (van Staa et al, 2014), Randomised Registry Trials (James et al, 2015), Platform Trials (Berry et al, 2015) and Trials within Administrative Databases (Anderson et al, 2016).

Researchers who implement new trial designs and create these new architectures attempt to take advantage of routine (non-experimental) data collection systems in order to improve recruitment, achieve economies of scale, reduce waste, and improve the generalizability of findings from trials. We describe how these new architectures aim to bring about economies of scale, either by 1) embedding multiple trials within one data collection and/or recruitment system (e.g. Trials within Cohorts, Platform Trials), or 2) embedding trials within a data collection system that has multiple purposes (e.g. Trials within Administrative Databases).

Some designs embed multiple trials within researcher generated data structures (Trials within Cohorts, Platform Trials). Others embed their trials within data structures with a routine healthcare purpose, e.g. Randomised Registry Trials (which use disease registers or registers of use of devices to identify patients and/or collect outcomes), Trials within Administrative Databases and Trials within Electronic Health Records. Almost all use a common comparator (usual/ standard care) for every trial.

An ongoing international collaboration is working to develop reporting standards for these new designs.

Implications

Although acceptable to patients, clinicians, ethics committees, policy makers and research funders, there are still a number of questions for these designs. Where (and how) in our healthcare systems do these designs might best fit? Which designs are the most efficient in terms of time, cost, and data quality? Should we now conduct trials of trial designs to compare these new approaches to standard pragmatic RCT designs?

Contributors

Merrick Zwarenstein
University of Western Ontario

Brett Thombs
McGill University

Linda Kwakkenbos
Radboud University

Ole Frobert
Orebro University

Ed Juszczak
Oxford University

A85

BURDEN AND CLINICAL IMPACT OF COMBINATION THERAPY EXPLORATION FOR NEW ANTICANCER DRUGS

BENJAMIN CARLISLE

MCGILL UNIVERSITY

Introduction: New cancer drugs are often advanced into combination therapy testing, in order to maximize their therapeutic impact and range of application. However, little is known about the scope of such efforts, the burdens for patients, or the amount of clinically actionable information gleaned by exploring combinations for newly approved drugs.

Methods: We captured all clinical trials of combination therapy involving all 11 anticancer drugs first licensed by FDA between 2005-2007 using a search of Medline and Embase. We extracted trial-level data on drug-related adverse events, patient numbers, trial outcomes and licensing or guideline decisions deriving from published trials.

Results: Our search captured 448 trials of 294 different unique drug-combinations, in 77 indications; 100 (22.3%) of trials were randomized. Trials enrolled 45,062 patients, and involved a minimum of 12,850 drug-related grade 3-5 serious adverse events. Only 4 combination regimens representing 9,446 exposed patients were ultimately approved by the FDA or led to a positive phase 3 trial result. Another 19 were incorporated as recommendations in NCCN guidelines.

Discussion: Among clinical trials of cancer drugs approved by the FDA, a large number of combination trials were launched after approval, resulting in a high degree of patient burden. However, the yield of clinically actionable information in the form of FDA labeling revisions or positive phase 3 trial results is very limited. This has implications for the justification of patient-subject enrolment in clinical trials, interpretation of trial results, and for policy regarding the launch of new trials of combination therapy in cancer.

Contributors

Jonathan Kimmelman
McGill University

A86

SAMPLE SIZE ISSUES FOR NON-INFERIORITY CLINICAL TRIALS WITH MULTIPLE PRIMARY ENDPOINTS

DANILA AZZOLINA

UNIVERSITY OF PADUA

Introduction

In most clinical trials the sample size is based on achieving adequate power for the primary endpoint; however, this is not always the case. In fact, the evaluation of a therapeutical intervention may require the use of multiple primary endpoints to provide a comprehensive picture of all its benefits over the symptoms of a multifaceted disease.

Despite availability of methods for handling multiple endpoints in clinical trials and the recent FDA draft guidance to industry on multiple endpoints in clinical trials, released in 2017, some challenges remain (i.e., confusion surrounding the terminology, testing for non-inferiority and superiority in a non-inferiority trial, lack of agreement on multiplicity corrections and choice of the most appropriate method).

In this work we address some of these challenges by discussing a sample size problem for a non-inferiority device trial with multiple endpoints.

Material and Methods

The motivating example is a parallel two-arm device trial in knee replacement surgery aimed to assess the non-inferiority of implant A (versus competitor) on the variation of muscle strength (measured in μV) during the extension movement, between the pre-operatory and 1 year follow-up evaluation, for 5 different muscles.

Two scenarios are hypothesized:

1. multiple co-primary endpoints approach, aimed to show non-inferiority for all the endpoints (intersection-union test, Berger 1982)

2. multiple primary endpoints in which non-inferiority holds for at least one endpoint (union-intersection testing strategy, Roy 1953).

In both scenarios, a range of effect sizes (0.3, 0.5 and 0.8), non-inferiority margins (from 10% to 40% of effect size) and correlations (0.5 and 0.8) between endpoints have been considered.

Sample size has been computed considering approach proposed by Sozu and colleagues (2011), adapted to a non-inferiority trial setting, to achieve a power of 0.8, given an overall type I error rate of 0.05.

Results

The sample size computed for primary endpoints scenarios is smaller than corresponding sample size obtained for co-primary endpoint scenarios. The needed sample size increases with decreasing effect size and correlation.

Across different scenarios, the sample size ranges from a minimum of 11 patients per arm (primary outcome setting, correlation = 0.8, effect size = 0.8, margin = 40% of effect), to a maximum of 277 (co-primary outcome setting, correlation = 0.5, effect size = 0.3, margin = 10% of effect).

Conclusion

In this sample size estimation problem, no adjustment is needed to control the type I error rate when the aim is to show the statistical significance of all the co-primary endpoints. Thus, the crucial issue is to determine the appropriate sample size for the trial while controlling the overall power and carefully considering the correlation between the endpoints.

As acknowledged by others (Cohen 1988; Althunian 2017), once the appropriate method has been chosen, it is important to tune the design parameters (i.e., effect, correlation and margin), on the basis of literature or elicitation of experts' opinions.

Contributors

Paola Berchialla
University of Turin

Ileana Baldi
University of Padua

Dario Gregori
University of Padua

A87

EXTENSION OF THE PRECIS-2 TOOL TO ASSIST IN PLANNING TRIALS OF MULTI-LEVEL INTERVENTIONS

WYNNE NORTON

NATIONAL CANCER INSTITUTE

Background

The PRagmatic Explanatory Continuum Indicator Summary-2 (PRECIS-2) tool—including published articles and an interactive website—helps researchers design trials that achieve a preplanned balance of pragmatic and explanatory characteristics. The 9 domains outlined in PRECIS-2 are mostly related to individual participant (usually patient) factors (e.g., eligibility, recruitment, primary outcome) along the pragmatic-explanatory continuum. While PRECIS-2 has been used to indicate pragmatic design in cluster randomised trials testing interventions in primary care, there are crucial differences in testing multi-level interventions at different hierarchical levels that affect patients, providers, and organizations. The aim of this study was to adapt the PRECIS-2 tool to assist researchers in designing multi-level intervention trials, with the overall objective of extending and complementing the existing tool for application to a wider range of trials.

Methods

We used an iterative process to identify, adapt, operationalize, and pilot test the PRECIS-2 domains for multi-level intervention trials. First, we three researchers held a series of one-hour conference calls to adapt PRECIS-2 to improve the domain descriptions for trials with multiple levels of intervention. Second, two researchers independently used the adapted PRECIS-2 domains to rate a multi-level intervention trial protocol on the pragmatic-explanatory continuum (1 = explanatory, 5 = pragmatic). Ratings were compared and discrepancies discussed; any questions or unresolved discrepancies were subsequently discussed among all three researchers. The adapted PRECIS-2 domains were revised, as needed, and a second iteration of coding was conducted by the two researchers on a different multi-level intervention trial protocol until ratings were consistent between coders and consensus was reached. Finally, four multi-level intervention trial protocols published in the journal Implementation Science were coded by two researchers and compiled as case studies, with detailed explanations and justifications for coding each of the adapted PRECIS-2 domains, consistent with the format followed in the original PRECIS-2 article (Loudon et al., 2015).

Discussion

We will present the proposed adaptation of the PRECIS-2 tool now including domains assessing multi-level factors (i.e., patient, provider, and organizational), illustrated with case examples, and discuss potential uses when designing trials to test multi-level interventions along the explanatory-pragmatic continuum.

Contributors

Kirsty Loudon
University of Stirling

Merrick Zwarenstein
Western University

A88

CONDUCTING HIGH-IMPACT, HIGH-QUALITY CLINICAL TRIALS: SITE STAFF TRAINING AS A COMPONENT OF QUALITY

TAYLOR SWANKIE

RTI INTERNATIONAL

Consistent implementation of study protocols across multiple sites is necessary to safeguard the integrity of data collected, and secure the public's confidence in results from the clinical trial. Frequent and standardized protocol-specific training is vital to preserving data integrity. Inadequate training on study protocols has the potential to create discrepancies in implementation, data collection, and documentation among clinical research staff within a site or across the multiple clinical sites. Equivalent training on study protocols mitigates the impact of differences in staffing experience, staff turnover, and variations in site workflow across the multiple sites. A well-managed, high impact clinical trial must value training not only to assure data integrity, but to ensure data quality.

Data coordinating centers (DCCs) have traditionally hosted in-person and on-site trainings. While face-to-face training is preferred among clinical research staff, it's not always practical to administer in-person training. In-person training can be time consuming, expensive, lengthy, and constrain DCC and clinical site resources. These limitations make it challenging to conduct frequent and standardized trainings on study protocols across the multiple sites. DCCs should consider other methods of training to complement face-to-face trainings and remain vigilant in monitoring multi-site clinical trials. Additionally, the DCC should have procedures and processes for training and retraining clinical site staff to facilitate continuity in study processes across multiple sites. These processes should not detract from the shared responsibility of the DCC and the clinical site of ensuring that clinical site staff are adequately trained and have a sufficient understanding of the protocol.

This presentation will describe the Pelvic Floor Disorder Network's (PFDN) processes and procedures for implementing cross-site training, documenting training completion, and certifying training comprehension. During an internal evaluation of processes, the PFDN identified staff training as an area for quality improvement. To ensure site staff are adequately and regularly trained, the PFDN created the Course Tracker web-based system to facilitate and manage trainings, as well as complement other PFDN training processes.

First implemented in March 2017, the PFDN Course Tracker has been used to administer courses for the study, Apical Suspension Repair for Vault Prolapse In a Three-Arm Randomized Trial Design: "ASPIRe". Over 40 clinical site staff have reviewed this e-learning module and taken the quiz. Since March 2017, about 5 new research staff have joined the clinical site staffs. As part of their onboarding, the new site staff reviewed the e-learning module and passed the quiz to demonstrate sufficient knowledge of the protocol before they could conduct protocol activities. For the DCC, the PFDN Course replaced the previous method of tracking training on several excel documents. The PFDN Course Tracker is an effective DCC tool for managing and tracking site staff training across 8 clinical sites in a network that has several active and ongoing studies.

Contributors

Peter Robbins
RTI International

Michael Ham
RTI International

Amanda Shaffer
RTI International

Carolyn Huitema
RTI International

Katrina Burson
RTI International

A89

COLLABORATIVE APPROACH TO CONDUCT USER TESTING BY NON-TECHNICAL RESEARCH STAFF

NISHA GROVER

GEORGE WASHINGTON UNIVERSITY

To maintain a robust and functioning data management system, it is ideal to have a dedicated team of testers who are knowledgeable in the software's functionality as well as the stakeholders' requirements. Testing commonly utilizes a combination of strategies including: unit testing (usually completed by the developer), integration testing, end-to-end (or complete system testing), and user acceptance testing (completed by the user). Each type of testing serves an important purpose in the overall maintenance of the data management system.

In most cases, academic coordinating centers are heavily staffed with research personnel but are not adequately funded to provide dedicated testers to conduct the different levels of testing. User acceptance and regression testing must be conducted by the research staff (e.g., coordinating center research assistant/coordinator) who have study obligations with assigned higher priority status than software testing. Additionally, research staff are often not well versed in types of testing, but are well versed in expected behavior as well as the aptitude and capabilities of the end user (i.e., clinical center or nurse coordinator). To ensure that each software update includes integration, user acceptance, and regression testing, the software development team must provide the research team with the necessary testing tools (e.g., testing notes in non-technical terms). However, this can be time consuming and cumbersome for software developers.

In order to alleviate the burden for the developers, the George Washington University Biostatistics Center (BSC) Software Development team adapted a collaborative model of software development to create testing procedures that are more "user friendly" to non-technical staff. Behavior Driven Development (BDD) and Acceptance Test Driven Development (ATDD) are methods of software development that allow technical and non-technical staff to collaborate and provide input on functionality and scope of software design. Using these collaborative models as a guide, the BSC Software Development team collaborated with a staff person who is knowledgeable in both the technical aspect of the data management system as well as the stakeholders' requirements to help create "user friendly" and non-technical testing notes and guides along with additional resources to help conduct an effective user acceptance and regression testing. Guides were created for each issue included in the software release along with detailed steps to complete the regression and user acceptance testing.

The collaborative model was implemented for two consecutive software updates and received positive feedback from the non-technical staff as well as BSC Primary Investigators.

Contributors

Alla Sapozhnikova
George Washington University

A90

OBJECTIVE INFERENCE TO ADJUST FOR BASELINE COVARIATES IN THE ANALYSIS OF RANDOMIZED CONTROLLED TRIALS FOR INCREASING EFFICIENCY

PAOLA BERCHIALLA

UNIVERSITY OF TORINO

Introduction. Baseline covariates impact the outcome in many randomized controlled trials (RCTs). In fact, it has been shown that models that adjust for baseline covariates can substantially improve the statistical power of the analysis when the covariates are strongly/moderately prognostic. While this is justified for continuous outcomes, for binary outcomes, which require non-linear models where covariate adjustment may change the magnitude of the treatment effect, the situation is subtler. This is well-recognized in the Guideline on adjustment for baseline covariates in clinical trials issued by the EMA in 2015, which requires to pre-specify the variables to be included in the primary analysis in the protocol, for avoiding the potential selection of the combination of covariates that may lead to the most favorable treatment effect estimate, especially in non-linear models.

Objective. The objective of this study is to investigate which methods of adjusting for baseline covariate in the analysis of RCTs maximize the statistical power while retaining the type I error rate as well as unbiased estimate of treatment effect.

Materials and Methods. Semiparametric Inference allows for the separation of treatment effect estimation from baseline covariate adjustment. This is the case of the Targeted maximum likelihood estimator (TMLE), which is a double robust estimator, the Inverse-Probability-of-Treatment Weighting (IPTW) estimator, the parametric G-computation and the Semiparametric Locally Efficient (SLE) estimator. These methods are compared in a simulation study and applied to re-analyse data from the Prospective Randomized Open label Blinded Endpoint (PROBE) study conducted to detect the effect of DPP-4 inhibitor sitagliptin with respect to conventional therapy in type-2 diabetes. Comparison of the results with those obtained using standard robust estimators as well as other data-adaptive model selection approaches is also carried out in terms of operating characteristics, bias in the treatment estimate and bias in standard error of treatment estimate.

Results. TMLE, IPTW, G-computation and SLE estimators provide considerable improvement over the unadjusted estimates. Covariates selection procedures lead to a slightly inflation of the type I errors for TMLE, IPTW and SLE estimators, which on the other hand have the largest statistical power. Instead, type I error are closer to the nominal rate for G-computation estimator. Model misspecification does not increase the bias estimates for TMLE, whereas the other not double robust estimators show an increase in standard errors of treatment estimate. For the PROBE trial reanalysis, the trend of the adjusted analyses is toward a higher p-value.

Conclusions. Adjusting for baseline covariates predictive of outcome in the analysis of RCTs can have more power than Intention-To-Treat based tests. However, for some classes of models, when the regression model is misspecified, inflated type I error and potential bias on treatment effect estimate may arise. Estimators that allow for separating the baseline covariate adjustment from the

treatment effect estimation can avoid the concern of potential bias for covariates post hoc selection retaining the focus on objective inference on treatment effect. Limitations of this study rely on the assumption of independent, identically distribution of data, which is not necessarily the case in RCT.

Contributors

Veronica Sciannameo
University of Torino

Daniele Bottigliengo
University of Padova

Corrado Lanera
University of Padova

Danila Azzolina
University of Padova

Ileana Baldi
University of Padova

A91

PREDICTIVE MODELING OF OUTCOMES IN A CLINICAL TRIAL WITH MISSING BASELINE DATA

SONIA DAVIS THOMAS

RTI INTERNATIONAL

Objective

We present a strategy using multiple imputation and LASSO to account for missing data when building a prediction model in the PROTECT pediatric ulcerative colitis study.

Background

Developing a prediction model of treatment response from baseline biological characteristics is often an important clinical trial aim. The problem of sporadic missing data across baseline parameters becomes magnified for parameter selection methods and must be addressed. The NIDDK-sponsored Predicting Response to Standardized Pediatric Colitis Therapy (PROTECT) clinical trial aimed to predict 1-year treatment response outcomes of remission and treatment failure for 428 children. Over 50 potential predictors from blood, stool, endoscopic exams, and rectal tissue samples had variable amounts of missing data.

Methods

Multiple imputation (MI) with fully conditional specification generated 100 imputed datasets. Least absolute shrinkage and selection operator (LASSO) logistic regression identified predictors within each imputation. We used a pre-specified rule to determine the final subset of predictors based on the 100 LASSO models. A parameter was selected if (a) the parameter was included in the LASSO model with maximum cross-validated (CV) area under the receiver operator characteristic curve (AUC) for 90% of the datasets, or (b) the parameter was included in the more parsimonious LASSO model with CV-AUC within 1 standard error of the maximum AUC for 50% of the datasets. We refined this selected subset by fitting a final model with MI logistic regression by evaluating for non-linearity and interactions, and removing any parameter with $p > 0.05$. Model estimates and prediction characteristics were combined across the 100 datasets with Rubin's rule. Goodness of fit was assessed by the distribution of the Hosmer-Lemeshow test across the imputations. The R glmnet package was used for LASSO, all other steps used SAS Version 9.4.

Results

Missing data ranged from 0% to 10%, but was higher for tissue (14%) and stool (44%) samples, and one lab parameter obtained from the medical record (26%). Our model selection method identified a parsimonious subset of uncorrelated parameters with reasonable prediction characteristics for primary and key secondary treatment outcomes. Odds ratio estimates from the final MI model were generally closer to 1 (weaker) than a comparison model on complete cases, providing an ad-hoc shrinkage correction.

Conclusions

We demonstrated an effective framework for selecting a parsimonious set of parameters with good prediction characteristics in the presence of missing data. We first identified a subset of parameters using pre-defined rules applied to LASSO logistic regression across multiple imputed datasets and then finalized the model using MI logistic regression. Benefits include the use of all available patient data, cross-validated parameter selection to reduce over-fitting, and a reduction in parameter over-estimation by multiple imputation.

Contributors

Jessie Wang
University of North Carolina

Nathan Gotman
University of North Carolina

Alison Marquis
University of North Carolina

Lee Denson
Cincinnati Children's Hospital Medical Center

Jeffrey Hyams
Connecticut Children's Medical Center

A92

COMBINING SAS AND LATEX FOR DSMB REPORT GENERATION: THE ENRGISE PILOT STUDY

WALTER AMBROSIUS

WAKE FOREST SCHOOL OF MEDICINE

One fundamental aspect of running a randomized controlled trial is preparation of reports for the Data and Safety Monitoring Board (DSMB). Preparation and collation of reports can be quite time consuming and is often a manual process requiring extra effort for merging reports, text responses, and formatting by multiple individuals. Day-to-day monitoring of the ENRGISE (ENabling Reduction of low-Grade Inflammation in SEniors) Pilot Study is mostly done using web-based dynamic reports written in SAS which are available to clinical site personnel. These reports include information on screening, randomization, retention, adherence, demographics, adverse events. When accessed on the web, these reports are presented to the user in a downloadable pdf file. Since the programming for many of the required DSMB reports is already completed in our web-based dynamic reports it seemed efficient to use the same SAS programs to both output to a pdf file for clinical site use and, using the SAS LaTeX tagset, to generate LaTeX code for the DSMB report. With minimal changes to our programs, we can then create a LaTeX wrapper to include the individual tables and figures (as pdf files which could also be generated using any other program, such as R) and combine into one overall report. Most of these reports are presented overall and by masked treatment arm.

Advantages of this approach include the ability to use the same SAS programs for reports for study personnel and the DSMB, simplicity of preparing open and closed reports, automatic generation of the table of contents and cross references, ability to include narrative text easily, almost instantaneous correction of small errors (compilation takes a second or two for a 180 page report), and time savings for the second and subsequent reports.

Disadvantages include needing to edit the headers of the component LaTeX reports (primarily the first time, subsequent times is mostly cut-and-paste) and additional time for preparation of the first report. In this presentation, we will provide code showing what we've done, compare our approach to alternatives, and show the results. We gratefully acknowledge the National Institute on Aging of the National Institutes of Health for funding ENRGISE (U01AG050499).

Contributors

Michael Walkup
Wake Forest School of Medicine

Cynthia Stowe
Wake Forest School of Medicine

Laura Lovato
Wake Forest School of Medicine

Daniel Beavers
Wake Forest School of Medicine

A93

EXPLORING THE POTENTIAL OF CLINICAL TRIAL DATA SHARING: THE SPRINT CHALLENGE

LAURA LOVATO

WAKE FOREST UNIVERSITY SCHOOL OF MEDICINE

Thoughtful, transparent systems for the responsible sharing of clinical trial data are important in maximizing the contribution of the participants, staff, and sponsors. Researchers have long sought to identify methods and timelines for sharing clinical trial data that balances the productivity of study investigators with the usefulness public data release.

SPRINT was a randomized, controlled, clinical trial in 9,361 participants aged ?50 years at increased risk for cardiovascular disease (CVD), excluding those with advanced kidney disease, diabetes or prior stroke, to intensive or standard systolic blood pressure goals of <120 vs. <140 mm Hg, respectively. Primary results reported in the New England Journal of Medicine (NEJM) on November 26, 20151, included follow-up through August 20, 2015, when intervention was stopped early for benefit.

The datasets used for analysis in that publication were made available on November 1, 2016 to individuals or groups who agreed to participate in the SPRINT Data Analysis Challenge. The Challenge was an effort led by the NEJM, with support from the National Institutes of Health and SPRINT investigators, to explore the potential benefits of rapid sharing of data from clinical trials. Individuals or groups were invited to analyze the dataset underlying the SPRINT article, and identify novel scientific or clinical findings. Among 200 qualifying teams, 143 entries were received. Entries were judged by a panel of experts on the basis of: utility of the findings to clinical medicine, originality and novelty of the findings, and quality and clarity of the methods used. All submissions were also open for crowd voting among the 16,000 persons following the SPRINT Challenge. Cash prizes were awarded, and winners were invited to present their results.

Experiences with the SPRINT Data Analysis Challenge will be discussed. The advantages and disadvantages of rapid data sharing will be explored, including but not limited to: productiveness of study investigators vs. those not familiar with the study; limited data release, or release in 'waves'; and the perspective of diverse viewpoints on this issue.

1SPRINT Research Group, Wright JT Jr, Williamson JD, et al. A Randomized Trial of Intensive versus Standard Blood-Pressure Control. N Engl J Med [Internet] 2015;373(22):2103–16.

(Funded by the National Institutes of Health; ClinicalTrials.gov number, NCT01206062.)

Contributors

Joni Snyder
National Heart Lung and Blood Institute, National Institutes of Health

Sean Coady
National Heart Lung and Blood Institute, National Institutes of Health

Pam Miller
New England Journal of Medicine

A94

MISSING DATA IN THE COMPONENTS OF A BINARY COMPOSITE ENDPOINT: IMPLICATIONS FOR THE COMPOSITE

ELAINE PASCOE

THE UNIVERSITY OF QUEENSLAND

Composite endpoints are increasingly used as primary endpoints in randomized controlled trials. They are sometimes used when it is not clear which of two or more endpoints are clinically most relevant. However, the primary motivation for the increasing use of composite endpoints is a practical one - reduction in the sample size and therefore cost of a trial. Composite endpoints are typically binary (measured as presence or absence of an event) based on two or more binary component endpoints. If there is an event on at least one of the components, then there is an event on the composite. The greater the number of components, the more likely there is to be an event on the composite; hence, their value in reducing sample size. There has been much discussion about the validity of composite endpoints (e.g., heterogeneity in the frequency of component events in response to a given treatment); however, little has been said about missing data on the components and implications for the composite. We describe the reasons for component missing data and implications for the composite of arteriovenous fistula vascular (AVF) access failure in the FAVOURED (Fish oils and Aspirin in Vascular access OUtcomes in REnal Disease) trial which investigated the effect of omega-3 fatty acids (fish oil) and aspirin use in patients with advanced kidney disease undergoing AVF creation for hemodialysis on the composite of AVF access failure at 12 months. There were three components to this composite endpoint (thrombosis, abandonment, failure to cannulate during an assessment period), each with differential amounts of missing data and for different reasons. We describe and report results for seven different scenarios for incomplete data on the components and discuss the implications for analysis and interpretation of the composite.

Contributors

Andrea Viecelli
The University of Queensland

Kevan Polkinghorne
Monash University

Liza Vergara
The University of Queensland

Peta Paul-Brent
The University of Queensland

David Johnson
The University of Queensland

A95

THE STANDARDIZED OUTCOMES IN NEPHROLOGY – HEMODIALYSIS (SONG-HD) INITIATIVE: ESTABLISHING A CORE OUTCOME SET FOR TRIALS IN PATIENTS ON HEMODIALYSIS

ANDREA VIECELLI

PRINCESS ALEXANDRA HOSPITAL

Introduction/objective: The Standardized Outcomes in Nephrology - Hemodialysis (SONG-HD) initiative aims to establish a core outcome set to be reported in all trials in hemodialysis based on the shared priorities of patients and health professionals.

Background: The prevalence of people with end-stage kidney disease in need of hemodialysis is steadily rising. Hemodialysis is a burdensome, time-consuming, and costly treatment that is associated with impaired quality of life and increased morbidity and mortality. In the past decade, more than 1500 reports of randomized trials in hemodialysis have been published, yet substantive improvements in patient outcomes remain to be seen. To some extent, this may be due to inconsistent and selective reporting of highly variable outcomes that are often of limited importance to patients and clinicians which limits the reliability and comparability of outcomes across trials for shared decision-making.

Methods: SONG-HD used an evidence-based consensus process. The five phases to establish core outcome domains included: a systematic review of outcomes reported in trials in hemodialysis; focus groups with nominal group technique with patients and caregivers to identify and prioritize outcomes, and describe reasons for their choices; multi-stakeholder interviews to elicit individual values and perspectives on outcomes; an international three-round online Delphi survey with patients, caregivers and health professionals (i.e. clinicians, nurses, allied health professionals, researchers, policy makers, and other relevant stakeholders with expertise in hemodialysis) to achieve consensus on critically important outcomes; and a consensus workshop to establish the core outcome set. The phases to establish the core outcome measures for each core outcome domain included: a systematic review of outcome measures reported in clinical trials; an international multi-stakeholder survey to rate and rank selected outcomes using a 9-point Likert scale and Best Worst Scale; consensus workshops to discuss measurement properties and feasibility aspects; and pilot and validation studies to evaluate and validate the identified core outcome measures.

Results: In total, 1376 patients, caregivers and health professionals from 73 countries participated in the consensus process to identify the four core outcome domains for hemodialysis: fatigue, vascular access, cardiovascular disease, and mortality. For fatigue, the impact of fatigue on life participation was the most critically important dimension and a 3-item questionnaire to assess fatigue is currently being validated. For vascular access, the function of a hemodialysis access was considered of most critical importance and defined as the need for an intervention/procedure to maintain the use of the vascular access for hemodialysis. For cardiovascular disease, myocardial infarction and sudden cardiac death were identified as the most critically important outcomes. Pilot

and validation studies are in progress to ensure that the core outcome measures are feasible and robust.

Conclusions: Consistent reporting of the core outcome set – fatigue, vascular access function, myocardial infarction, sudden cardiac death, and all-cause mortality – as a minimum in all trials in hemodialysis will improve the integrity, comparability, usability, and potential impact of trial-based evidence to inform decision-making in hemodialysis. This may ultimately lead to improved outcomes that are meaningful and important to patients and their clinicians.

Contributors

Allison Tong
Sydney School of Public Health, University of Sydney

Angela Ju
Sydney School of Public Health, University of Sydney

Emma O'Lone
Sydney School of Public Health, University of Sydney

Jonathan Craig
Sydney School of Public Health, University of Sydney

Carmel Hawley
Department of Nephrology, Princess Alexandra Hospital

A96

INTEGRATING NEUROIMAGING CENTRAL REPOSITORY AND TRACKING IN CLINICAL TRIAL MANAGEMENT SYSTEM

JAEMYUNG KIM

MEDICAL UNIVERSITY OF SOUTH CAROLINA

An integrated neuroimaging central repository and tracking module has been implemented into a web-based integrated electronic data capture (EDC) and clinical trial management system (CTMS) for the NIH Stroke Trial Network, in order to enable study subject neuroimaging file central repository, tracking and central reading.

Stroke trials often use centralized neuroimaging readings (e.g., infarct or hemorrhage size) to assess baseline disease severity, eligibility, and outcome. This complex procedure involves neuroimaging collection, registration, de-identification, transferring, and reading, so it is often one of the main causes for trial delay, especially when neuroimaging tracking and coordination are manually done.

A neuroimaging tracking and central repository module has been developed within an integrated EDC-CTMS system, with seven components: 1) Imaging Case Report Form (CRF) is posted along with other CRFs based on the study's data collection schedule. 2) Imaging tracking ID is generated, when the imaging CRF is submitted after passing all data validation checks. 3) Site uploads imaging file package in DICOM format to a designated imaging repository folder using File Transfer Protocol (FTP) with the Transport Layer Security (TLS) and the Secure Sockets Layer (SSL) cryptographic protocols. 4) An automated Quality Assurance (QA) program checks the uploaded imaging package, including Tracking ID and anonymization. 5) The tracking status of imaging CRF is updated and an email notification is triggered to the imaging coordinator for a manual QA step. If rejected, an automated email is triggered for the site to upload a new imaging file package and the corresponding CRF is flagged with a rule violation. If accepted, the imaging package is moved to central repository folder. 6) An authorized imaging central reviewer can enter their review results into imaging central review CRF. 7) A special data back-up and security policy is applied for this module.

This neuroimaging module is generically designed and therefore can be used for clinical trials requiring central imaging data review and central repository.

Contributors

Jean-Christopher Arnaud
Medical University of South Carolina

Wenle Zhao
Medical University of South Carolina

Contributors