Contemporary Clinical Trials Communications 18 (2020) 100561

Contents lists available at ScienceDirect

Contemporary Clinical Trials Communications

journal homepage: http://www.elsevier.com/locate/conctc

Determining a Bayesian predictive power stopping rule for futility in a non-inferiority trial with binary outcomes

Anna Heath [a,b,c,*], Martin Offringa [a,d,e], Petros Pechlivanoglou [a,d], Juan David Rios [a], Terry P. Klassen [f,g], Naveen Poonai [h,i], Eleanor Pullenayegum [a,b], on behalf KidsCAN PERC Innovative Paediatric Clinical Trials Team

[a] Child Health Evaluative Sciences, The Hospital for Sick Children, Toronto, Canada
[b] Division of Biostatistics, Dalla Lana School of Public Health, University of Toronto, Canada
[c] Department of Statistics, University College London, London, United Kingdom
[d] Institute of Health Policy, Management and Evaluation, University of Toronto, Canada
[e] Division of Neonatology, University of Toronto, Canada
[f] University of Manitoba, Winnipeg, Canada
[g] Children's Hospital Research Institute of Manitoba, Winnipeg, Canada
[h] Departments of Paediatrics, Internal Medicine, Epidemiology & Biostatistics, Schulich School of Medicine & Dentistry, Western University, Canada
[i] Children's Health Research Institute, London, Ontario, Canada

ARTICLE INFO

Keywords:
Non-inferiority trial
Bayesian predictive power
Stopping rule
Procedural sedation
Trial design

ABSTRACT

Background/Aims: Non-inferiority trials investigate whether a novel intervention, which typically has other benefits (i.e., cheaper or safer), has similar clinical effectiveness to currently available treatments. In situations where interim evidence in a non-inferiority trial suggests that the novel treatment is truly inferior, ethical concerns with continuing randomisation to the "inferior" intervention are raised. Thus, if interim data indicate that concluding non-inferiority at the end of the trial is unlikely, stopping for futility should be considered. To date, limited examples are available to guide the development of stopping rules for non-inferiority trials.
Methods: We used a Bayesian predictive power approach to develop a stopping rule for futility for a trial collecting binary outcomes. We evaluated the frequentist operating characteristics of the stopping rule to ensure control of the Type I and Type II error. Our case study is the Intranasal Ketamine for Procedural Sedation trial (INK trial), a non-inferiority trial designed to assess the sedative properties of ketamine administered using two alternative routes.
Results: We considered implementing our stopping rule after the INK trial enrols 140 patients out of 560. The trial would be stopped if 12 more patients experience a failure on the novel treatment compared to standard care. This trial has a type I error rate of 2.2% and a power of 80%.
Conclusions: Stopping for futility in non-inferiority trials reduces exposure to ineffective treatments and preserves resources for alternative research questions. Futility stopping rules based on Bayesian predictive power are easy to implement and align with trial aims.
Trial registration: ClinicalTrials.gov NCT02828566 July 11, 2016.

1. Introduction

Non-inferiority trials are an increasingly important, but often challenging, paradigm in which novel treatments are compared to active controls [1,2]. The active controls are typically the standard of care and the novel treatment is expected to maintain the same level of effectiveness but is preferred for other reasons, such as safety or ease of administration. For example, intravenous (IV) ketamine is used to sedate children with extremity fractures while they undergo a fracture reduction [3–5]. However, IV insertion is painful and - in young children -

Abbreviations: INK Trial, Intranasal Ketamine for Procedural Sedation trial; IV, Intravenous; IN, Intranasal; DSMB, Data Safety Monitoring Board.

* Corresponding author. Child Health Evaluative Sciences, The Hospital for Sick Children, Peter Gilgan Centre for Research and Learning, 686 Bay Street, 11th Floor – L4 East, Toronto, ON, M5G 0A4, Canada.
E-mail address: anna.heath@sickkids.ca (A. Heath).

https://doi.org/10.1016/j.conctc.2020.100561
Received 16 January 2020; Received in revised form 23 March 2020; Accepted 30 March 2020
Available online 8 April 2020

must be performed by skilled personnel [6]. To combat this, ketamine could be administered intranasally (IN), which would be preferable for patients [7]. Thus, the Intranasal Ketamine for Procedural Sedation (INK) trial was designed to assess whether IN ketamine is non-inferior to IV ketamine.

Ethically, it is important to monitor trials to ensure that patients are not exposed to unsafe treatments and unintended adverse events. Typically, this monitoring does not evaluate the primary efficacy outcome. However, interim analyses of efficacy have been highlighted as an important aspect of non-inferiority trials [8]: if, based on results at an interim analysis, it becomes unlikely to conclude non-inferiority, then trial participants are being needlessly exposed to a potentially less effective treatment. Despite this important ethical consideration, there are limited examples of stopping for futility in non-inferiority trials [8]. For example, a recent review highlighted that only 36% of 72 non-inferiority trials in oncology considered a formal interim analysis, meaning that these trials may have failed to protect patients from inferior treatments [9].

Several methods have been proposed to stop trials for futility [10], particularly using the concepts of conditional and predictive power [11, 12]. These methods stop a trial for futility if the probability of a statistically significant result from the completed trial is low at an interim analysis. Conditional power calculates the probability of a statistically significant result from the trial based on assumptions about the effect size and underlying event rate, usually based on the null and alternative hypotheses [11]. This approach has been criticised, as the chosen values for these quantities may not be appropriate in the face of the evidence. To address this concern, predictive power is the average conditional power over the current beliefs about the parameters of interest, usually determined using Bayesian methods [13].

These measures have been presented from a theoretical perspective [10–16] but have only been implemented in a small number of trials, e.g. Refs. [17–20], especially for non-inferiority trials [21]. In this paper, we develop a stopping rule for futility based on predictive power using the INK trial. We adjust the trial sample size to maintain power whilst also considering stopping for futility [22]. To encourage the development of stopping rules for futility in non-inferiority trials, we provide code to develop stopping rules, written in the R language for statistical computing [23]. We also provide a web application to implement this code without using R directly.

2. Methods

Spiegelhalter and others define predictive power as the chance of having a *positive result* from a trial, based on the *currently available data* [12]. Thus, predictive power can be calculated once we define a positive trial result and a method to analyse the currently available data that considers uncertainties in our knowledge about the effect size and underlying event rate in the population, known as the parameters of interest for our study.

Firstly, we define a positive trial result as a setting where the null hypothesis is rejected following the completion of the trial [13]. This implies that the trial would stop at the interim analysis point only when there is little chance of concluding non-inferiority following the completion of the trial. Secondly, to capture uncertainty in our knowledge about the parameters, we use Bayesian methods to analyse the data collected up to the interim analysis. These methods combine the interim data with a prior distribution that represents the beliefs of the researcher (s), from either expert opinion or previous studies, before undertaking the trial. By combining the prior distribution and the data, the predictive power takes into account all the available information about the parameters and formally accounts for our uncertainty when assessing whether the trial should be stopped [12].

2.1. A rejection region for binary outcomes

A successful trial result is defined as rejecting the null hypothesis following the completion of the trial. For a non-inferiority trial with binary outcomes, as in our example, the null hypothesis is

$$H_0 : p_C - p_N > \eta$$

where p_N and p_C are the probabilities of a favorable outcome for the novel treatment and the active control respectively and η is the non-inferiority margin. For the INK trial, p_N is the probability of experiencing adequate sedation for the duration of a fracture reduction using IN ketamine, p_C is the probability of experiencing adequate sedation for the duration of a fracture reduction using IV ketamine. The non-inferiority margin, η, was defined as 0.17 and was based on a survey we undertook of over 200 physicians. Note that this non-inferiority margin allows for a substantial drop in effectiveness. This is specific to the INK trial as IV ketamine can still be used when IN administration fails and therefore, patient care is minimally affected by this change in effectiveness.

To compute the probability of rejecting the null hypothesis, we must determine the "rejection region" of the hypothesis test, i.e. enumerate all the test statistics that would lead us to conclude non-inferiority at the end of trial. To compare two binary outcomes, we would compute the following statistic at the end of the trial:

$$T = \frac{N_C}{N} - \frac{N_N}{N}$$

where N_N and N_C are the number of patients who experience a "success" for the novel treatment and the active control respectively. The denominator N is the number of patients enrolled for each arm of the trial, which is assumed equal across the two arms, but this can be relaxed. Thus, the statistic T is the difference in the proportion of patients who have a successful outcome in the trial.

Based on the statistic T and a given trial sample size $2N$, we can determine the exact rejection region for the test of non-inferiority [24]. Initially, we calculate the probability of observing each possible value of N_C and N_N at the boundary of the null hypothesis using a binomial distribution. The probability of success for these binomial distributions are set by fixing a value of p_C and then fixing $p_N = p_C - \eta$. For example, in the INK trial, we used the estimate of $p_C = 0.97$, obtained from the literature [25].

Once we have specified every possible value for T and the associated probability of observing T under the null, we can determine the rejection region of the test. Specifically, we would reject the null hypothesis if T is less than a threshold value t^*. This is because the difference in proportions should be equal to or larger than the non-inferiority margin under the null. In practice, we determine the value of t^* by ordering the values of T, smallest to largest, and computing t^* as the largest value of T such that the probability of remaining below t^* is below α, the level of significance. As T can only take a finite number of values, it is not possible to fix the size at α and therefore we chose t^* as large as possible while the size of the test remains below α. As T and the associated probabilities are dependent on the trial sample size $2N$, the value of t^* will change for each sample size.

2.2. Bayesian analysis at interim

To undertake a Bayesian analysis at the interim analysis, we must specify our prior beliefs about p_C and p_N. For computational simplicity, we recommend choosing priors from the beta family of distributions as this is the conjugate distribution for the binomial trial outcomes [26]. It has also been suggested that stopping rules for futility should be developed based on "optimistic" prior distributions [27]. This is because there are limited data available at the interim analysis and, thus, the prior could have a substantial impact on the results at this interim stage.

A prior that strongly assumes that the novel treatment is non-inferior to the active control ensures that the trial would only stop if the results based on the data collected before the interim analysis overwhelmingly support a conclusion of inferiority for the novel treatment.

In the INK trial, we choose our optimistic priors based on the outcomes in a published trial. Specifically, the prior for p_C was based on a published trial in which 34 successful sedations were seen in 35 participants [25]. To center the prior on the proportion of successful sedations seen this trial, while inflating our uncertainty in the historical data [28], we set the prior for p_C as a beta distribution with parameters $17 = \frac{34}{2}$ and $0.5 = \frac{1}{2}$. For p_N, we define the prior, using the same process, as a beta distribution with parameters 20.5 and 3, based on three previous studies where 41 patients out of 47 were successfully sedated [28–30]. In the absence of data, these optimistic priors give an expected success rate from the trial of 0.798 and a probability of 0.82 that the IN administration is non-inferior to IV.

2.3. Probability of a successful trial

At the interim analysis, the predictive power stopping rule is based on the probability that we reject the null hypothesis following the completed trial, using the prior information and the currently available data. To calculate this probability, we need to determine the predictive distribution of the possible datasets that would be collected if the study were completed. Based on this, we can determine the probability that these datasets lie within the rejection region on the hypothesis test. This requires us to compute the statistic T for each predicted study dataset and determine whether T is in the rejection region. The trial should then be stopped for futility if the probability of rejecting the null hypothesis at the end of the trial is less than a given threshold. In this example, we set this threshold at 20% as this is recommended to trade off the risk of i) not stopping the trial when there is evidence of inferiority and ii) keeping the increase in sample size, required to recover the power lost by implementing the stopping rule, manageable [31].

In practice, it is easiest to undertake this prior predictive analysis using simulation. This can be achieved by generating S simulations from the posterior distributions for p_C and p_N, available analytically as we used conjugate distributions. For each $s = 1, \ldots, S$, the data that would be collected in the remainder of the trial is simulated from two binomial distributions,

$$Y_C^s \sim Bin\left(N-n,\ p_C^s\right),$$

$$Y_N^s \sim Bin\left(N-n,\ p_N^s\right),$$

where N is the proposed sample size of each arm in the completed trial, n is the sample size for each arm at the interim analysis and p_C^s and p_N^s are the s-th simulated values from the posterior distribution of p_C and p_N, respectively. These simulated data are then added to the data collected at the interim analysis and the statistic T is calculated to determine whether T is in the rejection region of the hypothesis test. The proportion of simulated values for T that are in the rejection region estimates the probability that the null hypothesis will be rejected upon completion of the trial.

2.4. Frequentist operating characteristics

Before implementing this stopping rule for futility, we must ensure that the type I and type II error rates of the trial are not compromised. To achieve this, we calculate the type I and type II error rates using simulation methods. Initially, we simulate the number of successes at the interim analysis for both treatments from

$$Y_C \sim Bin(n,\ p_C)$$

$$Y_N \sim Bin(n,\ p_N),$$

conditional on fixed values for p_C and p_N. Based on these values of Y_C and Y_N, we use the stopping rule outlined in the previous section to indicate whether the trial should stop for futility. If the simulated trial is stopped for futility, then it does not continue so the null hypothesis is not rejected and no further trial simulation is required. However, if this simulated trial does not stop for futility, then we continue by simulating the number of patients who are adequately sedated in the remainder of the trial from

$$Y_C^f \sim Bin(N-n,\ p_C),$$

$$Y_N^f \sim Bin(N-n,\ p_N).$$

Based on the simulated results from the completed trial, we then evaluate whether the null hypothesis is rejected. By simulating M potential trials, we compute the proportion of simulations for which the null hypothesis is rejected and, thus, the frequentist operating characteristics.

To determine the type I error, p_C and p_N should be fixed to values that are consistent with the null hypothesis, e.g. $p_C = 0.97$ and $p_N = 0.97 - 0.17 = 0.8$ for the INK trial. To determine the type II error, p_C and p_N should be set to values that are consistent with the alternative hypothesis, e.g. $p_C = 0.97$ and $p_N = 0.872$ for the INK trial.

Considering stopping for futility leads to a reduction in power. To counter this, we slightly increase of the trial sample size to ensure sufficient power for the trial. To determine the sample size that gives the correct power, we initially set the trial sample size using standard sample size calculations for a non-inferiority trial of two binary outcomes and calculate the power of a trial that considers stopping for futility. We then repeatedly increased the trial sample size by 1 and computed the power of a trial that considers stopping for futility with this full sample size until the power of the proposed trial was equal to, or larger than, β.

3. Results

3.1. Initial sample size calculation

The initial sample size calculation indicated that a sample size of 266 per arm would be sufficient to ensure a one-sided type I error rate below 2.5%, a power of 80% with $p_C = 0.97$, $p_N = 0.872$ and a non-inferiority margin of 0.17. In practice, even though binary outcomes are discrete, a sample size of 266 participants per arm gives a type I error rate of 2.5%.

Based on this sample size, all possible values of T and their probability under the null hypothesis are plotted in Fig. 1 with the rejection region highlighted in red. This indicates that the null hypothesis is rejected when the difference between the two proportions is sufficiently small to indicate that p_N is non-inferior to p_C. For a sample size of 266, the null hypothesis is rejected when T is less than $t^* = 0.117$.

3.2. Stopping rule for futility

For operational reasons determined by the Data Safety Monitoring Board (DSMB) schedule, the interim analysis to consider stopping for futility for the INK trial was proposed at 25% enrolment. We used $M = 500,000$ simulations to estimate the frequentist operating characteristics of the INK trial as this gives a higher than 95% chance that the type I error rate is estimated correctly to two significant figures. To ensure a power of 80% with $p_C = 0.97$ and $p_N = 0.872$, the sample size of the INK trial must increase to 280 per arm with $t^* = 0.118$. In this setting, the interim analysis should occur after 70 patients have been enrolled for each of the two treatment options, as 25% of 280 is 70. The type I error rate for this trial is 2.2%.

Fig. 2 displays the complete analysis of stopping rules for the INK trial as a function of the number of participants who are inadequately sedated, i.e., the participants who fail for the given treatment, out of the

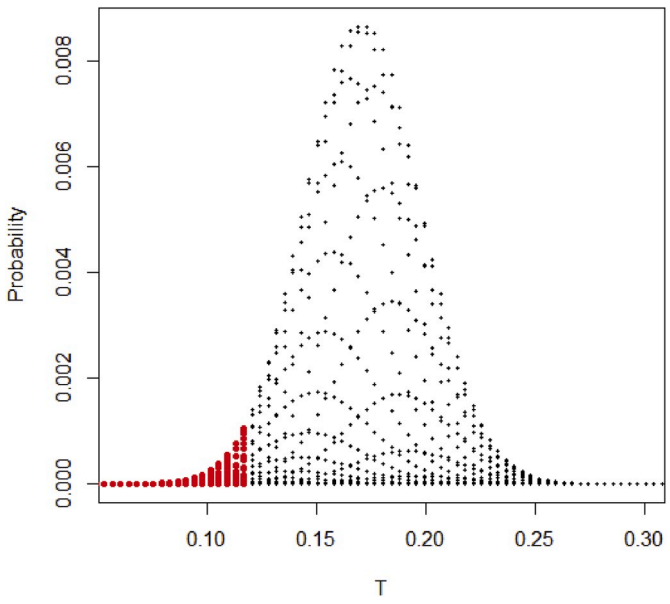

Fig. 1. The value of test statistic T plotted against the probability of observing T at the boundary of the null hypothesis. The bold dots (shown in red) represent values of T that are in the rejection region, i.e., the values of T that are less than $t^* = 0.115$. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

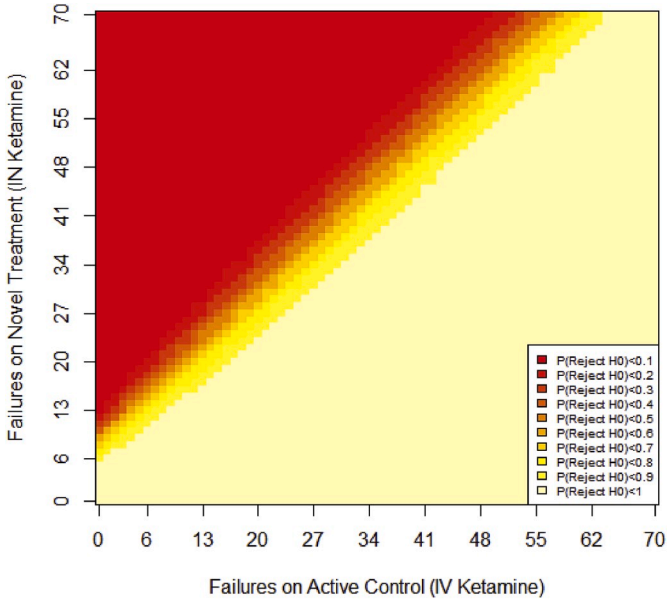

Fig. 2. The probability of rejecting the null hypothesis upon completion of the INK trial for different combination of failures on the active control (x-axis) and novel treatment (y-axis) at the interim analysis of the INK trial, proposed after an enrollment of 70 participants in each trial arm.

70 participants enrolled in each trial arm. These stopping rules are generated with $S = 5000$ simulations. In the darker the color, the lower the probability that the null hypothesis is rejected at the end of the trial. When more patients are inadequately sedated using IV ketamine (Active Control treatment failure), more patients can fail using the novel IN ketamine before the trial stops for futility. In addition, and by design, we can always experience a greater failure rate using IN ketamine compared to IV ketamine. Finally, if more sedation failures occur using IV ketamine then it is very likely that the null hypothesis will be rejected upon completion of the trial.

Most of the scenarios displayed in Fig. 2 are *very* unlikely to occur. For example, Fig. 2 highlights that the INK trial would not be stopped for futility if all patients were inadequately sedated with the active control, IV ketamine. However, as IV ketamine is assumed 97% effective, it is highly unlikely that all participants will be inadequately sedated. In Table 1, we summarize the most likely scenarios under which the INK trial would be stopped for futility. Table 1 also highlights the probability that each of these scenarios occurs, conditional on our prior beliefs about p_C and p_N. These probabilities are calculated based on 5000 simulations from the prior of p_C and p_N. These probabilities are low, highlighting that we have used optimistic prior distributions as current evidence indicates that the INK trial would not stop for futility. In total, the prior probability that the INK trial would stop for futility is 0.212. While this is relatively elevated, the chosen priors indicate an 18% chance that the Novel Treatment is non-inferior to the Active Control. So, crudely speaking, the majority of the time when the INK trial stops for futility, the Novel Treatment is truly inferior to the Active Control.

3.3. Power analysis

The power for the INK trial is highly dependant on the chosen values of p_N. Therefore, Table 2 displays power of the INK trial for different assumptions about the underlying success rate for the Novel Treatment. Note that these is significant drop in power as the true probability of success approaches the non-inferiority margin.

To facilitate the development of fuitility stopping rules, code in the statistical computing language R [23] is provided in the supplementary material to produce all the results in this manuscript. Additionally, a web application is available at http://annaheath.shinyapps.io/StoppingRulesFutilityEfficacy to reproduce these results without interfacing with R directly.

4. Discussion

It is important to consider stopping for futility in non-inferiority trials when there is little chance of concluding non-inferiority at the end of the trial as this prevents patients from receiving a potentially inferior treatment [8]. This article describes a Bayesian predictive power stopping rule for a non-inferiority trial with a binary primary outcome, using the INK trial as a case study. In this case study, implementing a Bayesian predictive power stopping rule for futility increases the sample size from 266 to 280 per arm to maintain 80% power. The stopping rule is expressed graphically and in a table as a function of the number of treatment failures observed in each trial arm to facilitate the presentation of results with key stakeholders, such as the Data Safety Monitoring Board (DSMB) and trial steering committee. We encourage the use of

Table 1
The stopping rule for the INK trial is based on the number of failed sedations observed for both the active control and the novel treatment at the interim analysis after 70 patients in each arm. Each row displays, conditional on the numbers of failed sedations for the active control (IV Ketamine), the minimum number of failures needed for novel treatment (IN ketamine) to stop the INK trial. The probability of stopping the trial is calculated as a sum, conditional on a fixed number of failures for the active control, across all possible number of failures for the novel treatment. The table displays the stopping rules that are relatively likely to occur at the interim analysis of the INK trial.

Number of failures observed on Active Control (IV Ketamine)	Minimum number of failures needed for Novel Treatment (IN Ketamine) to stop INK trial.	Probability of that the given scenario occurs
0	12	0.112
1	13	0.045
2	14	0.024
3	15	0.013
4	16	0.008
5	18	0.004

Table 2
The power of the INK trial with a stopping rule for futility for different assumptions about the probability of success for the Novel Treatment (p_N).

Probability of Success for Novel Treatment (p_N)	0.92	0.90	0.88	0.872	0.85	0.83
Power	0.998	0.98	0.88	0.80	0.46	0.19

stopping rules for futility in non-inferiority trials so resources, in terms of time and money, can be used for more promising research questions when there is little evidence that the trial could yield a change in clinical practice [32].

Previously, Bayesian predictive power methods have been computationally intense to use, which has limited their applicability within clinical trials [16]. However, we used conjugate distributions for binary outcomes to reduce the computational burden so the frequentist and Bayesian properties of the trial can be easily determined. This ensures that the type I and type II error rates of the trial are maintained when we consider stopping for futility. The main limitation of using conjugate distributions is that it places some restriction on the distributional form of the prior distributions for p_C and p_N, e.g. the prior distributions are independent. However, as the beta family of distributions is flexible, this restriction to conjugate distributions puts limited restrictions on the analysis. Thus, the methods presented in this paper can be implemented easily and efficiently across a range of studies to design stopping rules for futility.

Another limitation of this study is that the binomial distribution is a discrete distribution, which means that the trial cannot be run with the exact type I error rate of 2.5%. A discrete distribution, i.e. patients can either have a favorable or non-favorable outcome, means that there are a fixed number of potential values for T. Thus, the probability of observing $T < t^*$ is unlikely to be exactly equal to the proposed level of significance. For this study, we chose t^* as large as possible whilst remaining below 2.5% which caused the type I error rate for the INK trial with a futility stopping rule to equal 2.2% rather than 2.5%.

Going forward, the R functions and web application allow users to develop a stopping rule for their proposed trial by modifying the priors, the size and power of the underlying hypothesis test and the sample size at the interim analysis. Based on these user inputs, the provided code and interface will determine the sample size of the trial, calculate the stopping rules for futility and use simulation methods to compute the frequentist operating characteristics of the trial and the Bayesian probability of stopping the trial for futility. They also produce the graphical and tabular summaries presented in this article so the stopping rules can be more easily digested and implemented by the trial's DSMB and steering committee.

Conditional power has been proposed as an alternative method to develop stopping rules for futility that may yield alternative results [11,12]. Thus, future work should focus on comparing the properties of the stopping rule based on predictive power, presented in this article, with a stopping rule based on conditional power. This would allow for the development of code and a web application to further facilitate the development of stopping rules for futility in non-inferiority trials.

5. Conclusions

This article has presented a worked example to support the development of stopping rules for futility in non-inferiority trials and provided specialist, open-source software to support the design of new trials. Thus, Bayesian predictive power methods can now be used simply to create stopping rules for non-inferiority trials with binary outcomes.

Funding

The work is funded through an Innovative Clinical Trials Multi-year Grant from the Canadian Institutes of Health Research (funding reference number MYG-151207; 2017 - 2020), as part of the Strategy for Patient-Oriented Research and in partnership with the Alberta Children's Hospital Research Institute (Calgary, Alberta), Centre Hospitalier Universitaire Sainte-Justine (Montreal, Quebec), Children's Hospital Research Institute of Manitoba (Winnipeg, Manitoba), Children's Hospital of Eastern Ontario Research Institute (Ottawa, Ontario), Hospital for Sick Children Research Institute (Toronto, Ontario), Research Manitoba (Winnipeg, Manitoba), Department of Pediatrics, University of Western Ontario (London, Ontario), and the Women and Children's Health Research Institute (Edmonton, Alberta).

Data statement

Data were not used to produce the results in this article. All code used to produce the results has been included in the supplementary material.

Ethics approval and consent to participate/for publication

Not applicable.

CRediT author Statement

Anna Heath: Conceptualization, Methodology, Software, Formal Analysis, Writing – Original Draft, Visualisation.

Martin Offringa: Conceptualisation, Writing – Review and Edit, Supervision.

Petros Pechlivanoglou: Conceptualisation, Writing – Review and Edit, Supervision.

Juan David Rios: Conceptualisation, Writing – Review and Edit, Visualisation.

Terry P Klassen: Writing – Review and Edit, Funding Acquisition.

Naveen Poonai: Conceptualisation, Data Curation, Writing – Review and Edit.

Eleanor Pullenayegum: Conceptualisation, Methodology, Writing - Review & Editing, Supervision.

Declaration of competing interest

We have no conflicting interests to declare.

Acknowledgments

We thank all members of the KidsCAN PERC Innovative Pediatric Clinical Trials INK Study Team and for assistance in developing the INK trial protocol and statistical analysis plan. We would also like to acknowledge the KidsCAN PERC Innovative Pediatric Clinical Trials team and our parent advisors who provided valuable input on the study design and documents. We want to thank the Pediatric Emergency Research Canada (PERC) network of health care professionals and the KidsCAN Trials Network for their contribution and support to this project and pediatric clinical research in Canada.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.conctc.2020.100561.

References

[1] R. D'Agostino Sr, J. Massaro, L. Sullivan, Non-inferiority trials: design concepts and issues–the encounters of academic consultants in statistics, Stat. Med. 22 (2) (2003) 169–186.
[2] L. Mauri, R. D'Agostino Sr, Challenges in the design and interpretation of noninferiority trials, N. Engl. J. Med. 377 (14) (2017) 1357–1367.
[3] J. Cheng, W. Shen, Limb fracture pattern in different pediatric age groups: a study of 3,350 children, J. Orthop. Trauma 7 (1) (1993) 15–22.

[4] K. Jones, D. Weiner, The management of forearm fractures in children: a plea for conservatism, J. Pediatr. Orthop. 19 (6) (1999) 811.
[5] S. Schofield, J. Schutz, F. Babl, Procedural sedation and analgesia for reduction of distal forearm fractures in the paediatric emergency department: a clinical survey. Emerg. Med. Australasia (EMA) 25 (3) (2013) 241–247.
[6] E. Cummings, G. Reid, G. Finley, P. McGrath, J. Ritchie, Prevalence and source of pain in pediatric inpatients, Pain 68 (1) (1996) 25–31.
[7] A. Graudins, R. Meek, D. Egerton-Warburton, E. Oakley, R. Seith, The PICHFORK (Pain in Children Fentanyl or Ketamine) trial: a randomized controlled trial comparing intranasal ketamine and fentanyl for the relief of moderate to severe pain in children with limb injuries, Ann. emergancy med. 65 (3) (2015) 248–254.
[8] Z. Su, M. Stuntz, Futility stopping in non-inferiority trials, Contemp. Clin. Trials Commun. 13 (2019) 1–2.
[9] S. Tanaka, Y. Kinjo, Y. Kataoka, K. Yoshimura, S. Teramukai, Statistical issues and recommendations for noninferiority trials in oncology: a systematic review, Clin. Canc. Res. 18 (7) (2012) 1837–1847.
[10] P. He, T. Lai, O. Liao, Futility stopping in clinical trials, Stat. Interface 5 (2012) 415–423.
[11] J. Lachin, A review of methods for futility stopping based on conditional power, Stat. Med. 24 (18) (2005) 2747–2764.
[12] D. Spiegelhalter, L. Freedman, P. Blackburn, Monitoring clinical trials: conditional or predictive power? Contr. Clin. Trals 7 (1986) 8–17.
[13] S. Choi, P. Pepple, Monitoring clinical trials based on predictive probability of significance, Int. Biom. Soc. 45 (1) (1989) 317–323.
[14] A. Dmitrienko, M. Wang, Bayesian predictive approach to interim monitoring in clinical trials, Stat. Med. 25 (13) (2006) 2178–2195.
[15] D. Spiegelhalter, L. Freedman, A predictive approach to selecting the size of a clinical trial, based on subjective clinical opinion, Stat. Med. 5 (1986) 1–13.
[16] B. Saville, J. Connor, G. Ayers, J. Alvarez, The utility of Bayesian predictive probabilities for interim monitoring of clinical trials, Clin. Trials 11 (2014) 485–493.
[17] R. Lee, D. Lakkireddy, S. Mittal, C. Ellis, J. Connor, B. Saville, D. Wilber, Percutaneous alternative to the Maze procedure for the treatment of persistent or long-standing persistent atrial fibrillation (aMAZE trial): rationale and design, Am. Heart J. 170 (6) (2015) 1184–1194.
[18] J. Connor, K. Broglio, V. Durkalski, W. Meurer, K. Johnston, The Stroke Hyperglycemia Insulin Network Effort (SHINE) trial: an adaptive trial design case study, Trials 16 (1) (2015) 72.
[19] J. Dignam, J. Bryant, H. Wieand, B. Fisher, N. Wolmark, Early stopping of a clinical trial when there is evidence of no treatment benefit: protocol B-14 of the National Surgical Adjuvant Breast and Bowel Project, Contr. Clin. Trials 19 (6) (1998) 575–588.
[20] T. Julian, P. Blumencranz, K. Deck, P. Whitworth, D. Berry, S. Berry, A. Rosenberg, A. Chagpar, D. Reintgen, P. Beitsch, R. Simmons, Novel intraoperative molecular test for sentinel lymph node metastases in patients with early-stage breast cancer, J. Clin. Oncol. 26 (20) (2008) 3338–3345.
[21] W. White, D. Grady, L. Giudice, S. Berry, J. Zborowski, M. Snabes, A cardiovascular safety study of LibiGel (testosterone gel) in postmenopausal women with elevated cardiovascular risk and hypoactive sexual desire disorder, Am. Heart J. 163 (1) (2012) 27–33.
[22] S. Snapinn, M. Chen, Q. Jiang, T. Koutsoukos, Assessment of futility in clinical trials, Pharmaceut. Stat. 5 (2006) 273–281.
[23] R. R Core Team, A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2018.
[24] S. Wellek, Statistical methods for the analysisof two-arm non-inferiority trials with binary outcomes, Biom. J. 47 (1) (2005) 48–61.
[25] N. Kannikeswaran, M. Lieh-Lai, M. Malian, B. Wang, A. Farooqi, M. Roback, Optimal dosing of intravenous ketamine for procedural sedation in children in the ED—a randomized controlled trial, AJEM (Am. J. Emerg. Med.) 34 (8) (2016) 1347–1353.
[26] P. Lee, Bayesian Statistics, Arnold Publication, London, UK, 1997.
[27] D. Spiegelhalter, K. Abrams, J. Myles, Bayesian Approaches to Clinical Trials and Health-Care Evaluation, John Wiley & Sons, Malden MA, 2004.
[28] M. Ghajari, G. Ansari, A. Soleymani, S. Shayeghi, F. Ardakani, Comparison of oral and intranasal midazolam/ketamine sedation in 3-6-year-old uncooperative dental patients, J. Dent. Res. Dent. Clin. Dent. Prospects 9 (2) (2015) 61.
[29] P. Gharde, S. Chauhan, U. Kiran, Evaluation of efficacy of intranasal midazolam, ketamine and their mixture as premedication and its relation with bispectral index in children with tetralogy of fallot undergoing intracardiac repair, Ann. Card Anaesth. 9 (1) (2006) 25–30.
[30] D. Tsze, D. Steele, J. Machan, F. Akhlaghi, J. Linakis, Intranasal ketamine for procedural sedation in pediatric laceration repair: a preliminary report, Pediatr. Emerg. Care 28 (8) (2012) 767–770.
[31] J. Ware, J. Muller, E. Braunwald, The futility index: an approach to the cost-effective termination of randomized clinical trials, Am. J. Med. 78 (1982) 635–643.
[32] D. Schoenfeld, M. Meade, Pro/con clinical debate: it is acceptable to stop large multicentre randomized controlled trials at interim analysis for futility, Crit. Care 9 (1) (2004) 34.