UNIVERSITÀ DEGLI STUDI FIRENZE

UNIVERSITÀ DEGLI STUDI DI PERUGIA

INdAM Istituto Nazionale di Alta Matematica

Università di Firenze, Università di Perugia, INdAM consorziate nel CIAFM

DOTTORATO DI RICERCA
IN MATEMATICA, INFORMATICA, STATISTICA

CURRICULUM IN STATISTICA
CICLO XXXI

Sede amministrativa Università degli Studi di Firenze
Coordinatore Prof. Graziano Gentili

Bayesian HPD-based sample size determination using semi-parametric prior elicitation

Settore Scientifico Disciplinare MED/01

Dottoranda:
Danila Azzolina

Tutore
Prof. Ileana Baldi
Referente
Prof. Michela Baccini

Coordinatore
Prof. Graziano Gentili

Anni 2015/2018

A Nonna Virginia

che c'è stata e ci sarà sempre…

Vorrei ringraziare in primis la prof.ssa Ileana Baldi, relatore di questa tesi di laurea, per l'aiuto fornitomi, la conoscenza che mi ha donato, per la precisione, la dedizione e la pazienza dimostrate durante tutto il periodo di stesura e per il supporto scientifico fornito, in generale, in questi anni. Senza di lei questo lavoro non avrebbe preso vita!

Un ringraziamento speciale anche al prof. Dario Gregori, per le opportunità fornite che sono impagabili, per avermi dato la possibilità di affacciarmi al mondo della ricerca, quando ormai pensavo che fosse troppo tardi, per il fatto di aver creduto in me più di quanto non avessi fatto io stessa e per avermi dato la possibilità di imparare tante cose.

Vorrei ringraziare anche Paola Berchialla per i suoi preziosi consigli offerti e la disponibilità mostrata in questo percorso di ricerca.

Non posso non ringraziare il mio compagno Fabio per l'amore e la pazienza dimostrata e per aver avuto la costanza di sopportare la mia assenza e la mia totale dedizione al lavoro in questi anni.

Un grande ringraziamento a mia madre che con il suo dolce e instancabile sostegno, sia morale che materiale, mi ha permesso di arrivare fin qui davanti a voi oggi, contribuendo alla mia formazione personale.

Non posso non ringraziare le persone che ho incontrato questi anni. Corrado, amico instancabile, che è stato di supporto pratico e morale e fonte inesauribile di consigli soprattutto nei periodi difficili ma in tutto il percorso intrapreso in questi anni.

Non dimentico di ringraziare anche Giulia che è stata compagna di lavoro, e ha condiviso con me tante cose ed è stata reciproco sostegno nel superare le difficoltà che si sono presentate in questi anni. Ringrazio anche Daniele che mi ha dato forza in questi ultimi giorni con entusiasmo ed ottimismo.

Un sentito grazie a tutti!

Danila Azzolina

Padova, 29 ottobre 2018.

TABLE OF CONTENTS

Introduction

In several clinical trial settings, it is challenging to recruit the overall sample provided at the design stage. Several factors (i.e., high costs, regulatory barriers, narrow eligibility criteria and cultural attitudes towards research) can impact on the recruitment process [1]. The poor accrual problem is evident in the clinical research involving adults but also in the pediatric research, but also in pediatric research, where 37% of clinical trials terminate early due to inadequate accrual [2].

From a methodological-statistical standpoint, reduced sample size and the rarity of some diseases under consideration reduce a study's statistical power, compromising the ability to accurately answer the primary research question due to a reduction in the likelihood to detect a treatment effect [3]. This statistical point of view favors the use of a Bayesian approach to the analysis of clinical trial data. In recent years, Bayesian methods have increasingly been used in the design, monitoring, and analysis of clinical trials due to their flexibility [4].

In clinical trials candidate for early termination for poor accrual reasons, a Bayesian approach can incorporate the available knowledge provided by literature (objective prior) or by elicitation of experts' opinions (subjective prior) on the treatment effect under investigation [5] in order to reduce uncertainty in treatment effect estimation.

The first article (Chapter 1) shows the potentiality of the Bayesian method for use in pediatric research, demonstrating the possibility to include, in the final inference, prior information and trial data, especially when the small sample size is available to estimate the treatment effect. Moreover, this study aims to underline the importance of a sensitivity analysis conducted on prior definitions in order to investigate the stability of inferential conclusions concerning the different prior choices.

In a research setting where objective data to derive prior distribution are not available, an informative inference complemented with an expert elicitation procedure can be used to translate into prior probability distribution (elicitation) the available expert knowledge about treatment effect [5,6] The elicitation process in the Bayesian inference can quantify the presence of uncertainty in treatment effect belief. Additionally, this information can be used to plan a study design, e.g., the sample size calculations [7] and interim analysis [8]. Elicitation may be conducted in a parametric setting, assuming that experts' opinion may be represented by a good note family of probability distributions identified by hyper-parameters [6], or in a not parametric and semiparametric hybrid setting [9]. It is widely assessed that the primary limit of a parametric approach is to constrain expert belief into a pre-specified family distribution [10].

The second article (Chapter 2) aims to investigate the state-of-art of the Bayesian prior elicitation methods in clinical trial research performing an in-depth analysis of the discrepancy between the approaches available in the statistical literature and the elicitation procedures currently applied within the clinical trial research.

A Bayesian approach to clinical trial data may be defined before the start of the study, by the protocol, defining a sample size taking into account of expert opinion providing the possibility to use also nonparametric approaches. A more flexible sample size method may be suitable, for example, to design a study conducted on small sample sizes as a Phase II clinical trial, which is generally one sample, single stage in which accrued patients, are treated, and are then observed for a possible response [11].

Generally, Bayesian methods, available in the literature to obtain a sample size estimation for binary data, are based on parametric Beta-binomial solutions, considering an inference performed in term of posterior Highest Posterior Density interval (HPD) [12]. The aim of the third article (Chapter 3) is to extend the main criteria adopted for the Bayesian Sample size estimation, Average Coverage Criterion (ACC), Average Length Criterion (ALC) and Worse Outcome Criterion (WOC), proposing a sample size estimation method which includes also prior defined in a semiparametric approach to the prior elicitation of the expert's opinion [9]. In the research article also a practical application of the method to a Phase II clinical trial study design has been reported.

The semiparametric solution adopted is a very flexible considering a prior distribution obtained as a balanced optimization of a weighed sum two components; one is a linear combination of B-Spline adapted among expert's quantiles, another one is an uninformative prior distribution.

References

[1] Kadam RA, Borde SU, Madas SA, Salvi SS, Limaye SS. Challenges in recruitment and retention of clinical trial subjects. Perspectives in Clinical Research 2016;7:137.

[2] Pica N, Bourgeois F. Discontinuation and nonpublication of randomized clinical trials conducted in children. Pediatrics 2016:e20160223.

[3] Billingham L, Malottki K, Steven N. Small sample sizes in clinical trials: a statistician's perspective. Clinical Investigation 2012;2:655–7.

[4] Baiardi P, Giaquinto C, Girotto S, Manfredi C, Ceci A. Innovative study design for paediatric clinical trials. European Journal of Clinical Pharmacology 2011;67:109–15.

[5] O'Hagan A. Eliciting expert beliefs in substantial practical applications (Disc: p55-68). Journal of the Royal Statistical Society, Series D: The Statistician 1998;47:21–35.

[6] Garthwaite PH, Kadane JB, O'Hagan A. Statistical methods for eliciting probability distributions. Journal of the American Statistical Association 2005;100:680–701.

[7] Spiegelhalter DJ, Freedman LS. A predictive approach to selecting the size of a clinical trial, based on subjective clinical opinion. Statistics in Medicine 1986;5:1–13.

[8] Spiegelhalter DJ. Incorporating Bayesian ideas into health-care evaluation. Statistical Science 2004;19:156–74.

[9] Bornkamp B, Ickstadt K. A note on B-splines for semiparametric elicitation. The American Statistician 2009;63:373–7.

[10] Oakley JE, O'Hagan A. Uncertainty in prior elicitations: a nonparametric approach. Biometrika 2007;94:427–41.

[11] Kearney JM. Biostatistics in Clinical Trials. British Journal of Biomedical Science 2003;60:57.

[12] Joseph L, Wolfson DB, Du Berger R. Sample size calculations for binomial proportions via highest posterior density intervals. The Statistician 1995:143–54.

Chapter 1
Bayesian methods to cope with poor accrual in pediatric trials[1]

Danila Azzolina[1,2], Dario Gregori[1], Paola Berchialla[3], Ileana Baldi[1]…[2]

[1] Unit of Biostatistics, Epidemiology and Public Health, Department of Cardiac Thoracic Vascular Sciences and Public Health, University of Padova, Italy

[2] Department of Statistics, Computing, Applications "Giuseppe Parenti", University of Florence

[3] Department of Clinical and Biological Sciences, University of Torino, Italy

Corresponding author:

Ileana Baldi: Unit of Biostatistics, Epidemiology and Public Health, Department of Cardiac Thoracic Vascular Sciences and Public Health, Via Loredan 18, 35121 Padova, E-mail: ileana.baldi@unipd.it

[1] The article will be submitted to BMC Trial.

[2] The article is under review by the clinical team involved in the RESCUE trial, which provided us with the data and this interesting research insight. Their names will be added before submission.

Abstract

In several clinical trial settings, it is challenging to recruit the overall sample provided at the design stage. Several problems may occur in the patients' enrollment, and the amount of information conveyed by a trial terminated prematurely for poor accrual may be minimal. A Bayesian analysis of such trials may salvage this information by providing a framework in which to combine prior evidence with current evidence.

A Bayesian analysis is applied to a trial candidate terminated due to poor accrual to frame the conditions in which this approach can be used in practice, which is in line with the requirements of regulatory agencies. The randomized, double-blind RESCUE trial (EudraCT Number: 2013-000388-10), which compares the effect of adjunctive oral steroids on preventing renal scarring in infants with febrile urinary tract infections, is the motivating example.

This study aims to draw attention to the importance of a sensitivity analysis conducted on prior choices and to evaluate the robustness of inferential conclusions with respect to prior specifications. Different scenarios on prior definitions have been considered, defining different levels of penalization (discounting) on the historical information derived from other studies: 1) power prior without discounting (informative); 2) power prior with 50% discounting (low informative); and 3) power prior with 100% discounting (uninformative) corresponding to a beta (1,1) prior.

The results are compared in terms of the posterior probability for the absolute value of the difference in proportion, defined as $|\delta|$, evaluating the probability that $|\delta|$ may assume values greater than a threshold identified as 0.2 and 0.1.

$P(|\delta| > 0.1)$ is 0.93 in the formative scenario, 0.87 in the low-informative prior scenario and 0.64 considering an uninformative prior scenario. For $P(|\delta| > 0.2$, the values are 0.73, 0.66, and 0.33 in the informative, low informative and uninformative settings, respectively.

The results are more similar across scenarios with a margin of 0.1 compared to the results with a margin of 0.2. Assuming that $|\delta| = \delta$, the trial results seem not to be promising, indicating weak evidence in favor of the treatment and observing a likely absolute value of the effect of approximately 0.1. However, if the design provided by the protocol had anticipated an adaptive strategy, a decline in the futility scenario probably would have been one of the possible results of the study.

Introduction

Difficulties in the enrollment of the overall trial sample size, as indicated at the design stage, may be caused by a number of factors (i.e., high costs, regulatory barriers, narrow eligibility criteria and cultural attitudes towards research), which may challenge different research fields to different extents, depending on the characteristics of the population and the intervention under evaluation [1]. Prior research evaluating the reasons for termination across a broad range of trials, as registered in ClinicalTrials.gov, reports that insufficient enrollment is the most common reason, with a prevalence ranging from 33.7% to 57% depending on the definition used [2,3]. The poor accrual problem is marked not only in clinical research with adults, primarily in oncology [4–6] and cardiology [7], but also in pediatric research, where 37% of clinical trials terminate early due to inadequate accrual [8]. It is recognized that research conducted on children poses several methodological and ethical challenges [9]. Moreover, it is essential to consider that managing and conducting pediatric trials is more complicated than adult trials in terms of practical, ethical and methodological problems [10].

From a practical point of view, this kind of study may be initiated once a favorable benefit-risk balance for the treatment under investigation in adult patients has been assessed. For this reason, the FDA (US Food and Drug Administration) and EMA (European Medicines Agency) during the adult trial conduction require specific plans to be approved for experimentation to be conducted in pediatric patients [9,11]. Ethical-regulatory issues also inform this research framework. The conduction of pediatric trials is challenging because the studies are conducted on a vulnerable population and oftentimes on a limited number of patients. Moreover, the withdrawal of consent is a very sensitive issue in this research setting [12]. In this context, small sample sizes and ethical complications related to the acceptability of a study cause difficulties with patient enrollment. From a methodological-statistical standpoint, a reduced sample size and the rarity of some diseases under consideration reduce a study's statistical power, compromising the ability to accurately answer the primary research question due to a reduction in the likelihood to detect a treatment effect [13]. The scientific community has conveyed that early termination of a trial due to poor accrual creates inefficiency in clinical research, with consequent increases in costs [14] and a waste of resources as well as a waste of the efforts of the children involved in the experimentation [12].

For these reasons, alternative and innovative approaches to pediatric clinical trial design are a recent object of debate in the scientific community [9,11]. Alternative methods to pediatric trial design and analysis are also suggested in the ICH Topic E11 guidelines [15] as well as guidance for trial planning and design in pediatric context and in the guidelines that the EMA published a few years ago [16–18].

Data from trials terminated prematurely for poor accrual, however, can provide useful information for reducing the uncertainty about treatment effect. This statistical point of view favors the use of a Bayesian approach to the analysis of clinical trial data [13]. In recent years, Bayesian methods have increasingly been used in the design, monitoring, and analysis of clinical trials due to their flexibility [19]. Considering the research setting described in this work, the Bayesian methods used for accrual monitoring are also interesting [20]. These methods are well suited to designing and analyzing studies conducted with a small sample size and are particularly appropriate for studies involving children, even in cases of rare disease outcomes [9]. In clinical trials with candidates that terminated early for poor accrual reasons, a Bayesian approach may be useful for incorporating the available knowledge on the investigated treatment effect, as available via the literature or elicitation of experts' opinions [21].

Moreover, in a Bayesian setting, prior information combined with data may support the final inference for a trial conducted on a limited number of enrolled patients [22,23]. In pediatric trials, for example, the awareness that a treatment is effective in adults increases the probability of its efficacy in children. This awareness may be quantitatively translated into a prior probability distribution [9,11]. However, when there is a small sample size, the final inference may be severely conditioned by a misleading prior definition [23]. In this framework, the FDA suggests performing a sensitivity analysis on prior definitions [24], especially for very small sample sizes [25].

The primary purpose of this work is to show the potential for a Bayesian approach in pediatric research, demonstrating the possibility to include, in the final inference, prior information and trial data, especially when only a small sample size is available to estimate the treatment effect. This study also highlights the importance of conducting a sensitivity analysis on prior choices to evaluate the robustness of inferential conclusions with respect to the prior specification. The RESCUE (REnal SCarring Urinary infEction) trial, a pediatric trial candidate for early termination due to under-recruitment, serves as the motivating example.

Methods

RESCUE trial

The RESCUE (REnal SCarring Urinary infEction) trial was a randomized controlled double-blind trial. The purpose of this study was to evaluate the effect of adjunctive oral steroids to prevent renal scarring in young children and infants with febrile urinary tract infections. The primary outcome was the difference in scarring proportions between treatment arms. The study had been designed in a frequentist setting. According to the protocol, a sample size of 92 randomized patients per arm was required, which also considered 20% of losses to follow-up.

This case study illustrates the conceptual framework for Bayesian analysis, but it does not aim to assess the efficacy of the experimental treatment. For this reason, the two study arms will be undisclosed and labeled as arm A and arm B. After two years, only 17 recruited patients completed the follow-up for the study outcome (6 in arm B and 11 in arm A) involving a loss of the final power of 63%.

The main issues driving under-recruitment were:

1. **Procedural.** According to the study design, those that were eligible to be included in the study were patients aged between 12 and 24 months, those at high risk of renal scarring after urinary infection, febrile disease resulting in positive blood and catheter examination (PCT>1 ng/ml, and GB>=1 or nitrite positive).

   In clinical practice, catheter examination is considered more invasive than urine bag examination; consequently, of 339 first level eligible patients, only 81 were positive for catheter examination.
2. **Poor adherence of parents to the therapy**. Only 48 parents signed the informed consent form because the experimental treatment was considered unsafe for children's health by parents.
3. **Problems with the final diagnosis.** According to the study protocol, final scintigraphy was required to assess the presence of renal scarring. However, during the conduction of the study, the parents, after the resolution of acute febrile disease, thought that the final scintigraphy was useless.

Statistical Analysis

A beta-binomial model was used to analyze the difference in scar proportions between arms [26].

The two key ingredients were a prior distribution in the beta family to capture the background knowledge on the expected proportions and the binomial likelihood function to express the observed evidence emerging from trial data. The posterior distribution balances prior information with trial data.

This model is widely adopted in the Bayesian setting since it allows acquisition of a posterior conjugate distribution in closed form without problems related to the convergence of numerical integration [27]. The posterior distribution for the difference in proportions requires the estimation of the posterior distribution of the scar proportion in each arm, separately, and has been computed with the following resampling procedure [28]:

1. Resampling of the proportion of scarring π_A^* from $\pi_A|X_A$, which is the posterior distribution for arm A;

2. Resampling of π_B^* from $\pi_B|X_B$, which is the posterior distribution for arm B; and
3. Obtaining the posterior distribution for the parameters related to the difference in proportions between arms by calculating δ^* as the difference in proportions from the previously resampled distributions [29].

Resampling procedures were performed using an MCMC estimation algorithm, as indicated in the literature [28], using 3 chains, 5000 iterations, and 1000 adaptations. Computations were performed using OpenBUGS [30] and R version 3.3.2 [31].

Sensitivity Analysis

The inference was expected to be seriously conditioned by the prior choice, as only a few data points were available to estimate the likelihood. For this reason, a sensitivity analysis was performed to assess the robustness of the inferential conclusion with respect to the different prior choices.

In a Bayesian context, the prior reflects existing knowledge about the parameters of interest (in this case, the scar proportion) before observing the current trial data. The more prior information that is added, the more informative it becomes. Informative priors are expected to impact final inference through smaller credible intervals. Therefore, it may be suitable to control the impact of this information on the final analysis.

In some cases, a penalization defined for the prior distribution may reduce this impact. This procedure may also be required by a regulatory agency, such as the FDA [32]. Different levels of penalization (discounting) may be provided for the historical information using a power prior approach [33] in order to perform a sensitivity analysis on the prior choices. The historical information may be included in the final inference using a $Beta(\alpha_1, \beta_1)$ prior, where:

$$\alpha_1 = \alpha_0 d_0 + 1$$

$$\beta_1 = \beta_0 d_0 + 1$$

The α_0 and $\beta_0$values are the parameters defined by the number of successes and failures derived from the literature and are $(\alpha_0 - 1)$ and $(\beta_0 - 1)$, respectively. The value d_0 defines the amount of historical information to be included in the final inference. The discounting factor is otherwise defined as $(1 - d_0) \times 100$ and represents the levels of penalization (discounting) on the historical information derived from other studies.

1. If $d_0 = 0$, the data provided by the literature are not considered, indicating a 100% discounting on the historical information. According to this scenario, the prior is an uninformative $Beta(1,1)$ distribution.

2. If d_0=1, all the information provided by the literature is considered in the inference, indicating a 0% discounting on the historical data.

In this general setting, three different scenarios were hypothesized for the prior computation (Figure 1):

- **Power Prior without discounting (Informative, d_0=1).** A Beta informative prior was derived considering the number of successes and failures found in the literature [34], defining prior probability distributions as a $Beta(6,12)$ and a $Beta(39,26)$.
- **Power Prior 50% discounting (Low Informative, d_0=0.5).** The Beta prior with a 50% discount, defined in the literature as a Substantial-Moderate discounting factor [35], was defined on Beta parameters specified in the informative scenario. The discount procedure leads to control the effect of prior information on final inference as indicated in the literature [27].
- **Power Prior 100% discounting (Uninformative, d_0=0).** A $Beta(1,1)$ prior.

The posterior distribution was analyzed by considering the hypothesis that the absolute value of the difference in scar proportions $|\delta|$ is greater than two clinically meaningful effects of at least 0.1 and 0.2 in each scenario. The correct analysis procedure is to express the treatment effect as a difference in proportions δ. However, the absolute value of this difference $|\delta|$ was presented to mask the true treatment effect.

Results

Figure 2 shows the posterior distributions for $|\delta|$. The probability that $|\delta|$ is greater than 0.1 was very similar in the informative (0.93) and low informative (0.87) prior scenarios, while this probability was smaller for the uninformative scenario (0.64). When we consider a higher effect of 0.2, the results markedly differ across scenarios. Specifically, the probabilities of $|\delta|$ greater than 0.2 were 0.73, 0.66, and 0.33 in the informative, low-informative and uninformative setting, respectively.

Table 1 reports the 95% credibility intervals for the median $|\delta|$; the interval includes 0.1 for the estimates provided with a 50% discount in the uninformative prior setting and for the informative credibility interval. All the credibility intervals were larger than 40%. The informative prior results were in the smallest interval, with a posterior length of 0.41.

Discussion

Regulatory agencies advocate for an increase in pediatric research, which is motivated by the need for more information on treatment labeling to guide pediatricians and to offer more suitable and safe treatments to children [11]. However, in various cases, pediatric trials have demonstrated difficulties in enrolling participants [36].

The RESCUE trial represents a typical example of a complex trial in pediatric research, seriously conditioned by poor accrual. The difficulties encountered in the enrollment and retention of participants were complexity in the study protocol (i.e., catheter urine examination and renal scintigraphy [36,37]) and poor adherence to the therapy under consideration by the parents. The Bayesian analysis conducted on the RESCUE trial allowed investigators to combine information provided by current trial data (n = 17) with the evidence provided by the literature using the posterior distribution of the difference.

For the sensitivity analysis, the posterior distribution of the absolute difference in scar proportions was highly influenced by the prior choices and was weakly influenced by the data when using informative priors. In the uninformative prior setting, the inference seemed to be guided by the data, but the sample size was too small to return strong evidence. A discounting factor placed on prior parameters seems to be a good compromise to control the weight of prior on final inference, as well as taking into account the data [38].

In pediatric research, the optimal amount of discounting factor on an informative prior, an aspect limiting the influence of the prior distribution on final inference, remains to be discussed [11]. However, in the Bayesian inference generally, the weight given to the prior is associated with the subjective confidence of the importance of the prior distribution on the final result [38]. To perform sensitivity analysis on priors (i.e., to define the robustness of conclusions that may be affected by decisions made on the priors) is highly recommended also for pediatric trials [11]. This is in-line with the literature [23] and FDA recommendations [24].

The analysis of the RESCUE trial suggests that the absolute value of the effect is likely approximately 0.1. Assuming that the inference performed on the absolute value of the difference in proportion is comparable to the inferential conclusion obtained (considering a delta outcome), it is possible to assess that the trial results are not promising, probably outlining that the futility scenario is one of the possible results of the trial. However, in this work choosing a Bayesian design is advocated beforehand--eventually providing an adaptive strategy with an interim assessment for futility--not switching to a Bayesian analysis method that produces a more favorable outcome after observing the data.

Considering the clinical trial conducted in a similar research setting, a Bayesian analysis of trial data may represent an alternative methodological approach useful in analyzing data as suggested in the literature [9,11]. Of note, the choice of prior moderately affects this conclusion. We do not claim that this evidence should be regarded as definitive, but it can be reused to derive parameters on priors to be considered for analysis in other similar studies, as suggested by FDA guidance [39].

Study Limitations

The study was conducted considering only the conjugate prior Beta setting. It may be interesting also to explore the impact of inference in the case of posterior that is obtained not in closed form.
For example, instead of directly placing a parameter derived on Beta prior, it may be advisable to consider expert elicitation about treatment effects to define specific prior distribution. The overall research conclusions may be generalized, considered for future research development, and used to perform a specific simulation study.

Conclusion

Bayesian inference is a flexible tool compared to the frequentist one, especially for trials conducted in a poor accrual setting. A poor accrual setting often occurs in research conducted in the pediatric field, and Bayesian inference can consider prior knowledge about treatment effect, which is especially useful in such trials. However, we also have to take into account that data may only weakly influence the informative inference conducted on small samples. For this reason, sensitivity analyses of prior distributions allows for the evaluation of the robustness of inferential conclusions.

References

[1] Kadam RA, Borde SU, Madas SA, Salvi SS, Limaye SS. Challenges in recruitment and retention of clinical trial subjects. Perspectives in Clinical Research 2016;7:137.

[2] Pak TR, Rodriguez M, Roth FP. Why clinical trials are terminated. BioRxiv 2015:021543.

[3] Williams RJ, Tse T, DiPiazza K, Zarin DA. Terminated trials in the ClinicalTrials. gov results database: evaluation of availability of primary outcome data and reasons for termination. PLoS One 2015;10:e0127242.

[4] Rimel B. Clinical Trial Accrual: Obstacles and Opportunities. Frontiers in Oncology 2016;6:103.

[5] Mannel RS, Moore K. Research: An event or an environment? Gynecologic Oncology n.d.;134:441–2. doi:10.1016/j.ygyno.2014.08.001.

[6] Stensland KD, McBride RB, Latif A, Wisnivesky J, Hendricks R, Roper N, et al. Adult cancer clinical trials that fail to complete: an epidemic? JNCI: Journal of the National Cancer Institute 2014;106.

[7] Baldi I, Lanera C, Berchialla P, Gregori D. Early termination of cardiovascular trials as a consequence of poor accrual: analysis of ClinicalTrials. gov 2006–2015. BMJ Open 2017;7:e013482.

[8] Pica N, Bourgeois F. Discontinuation and nonpublication of randomized clinical trials conducted in children. Pediatrics 2016:e20160223.

[9] Baiardi P, Giaquinto C, Girotto S, Manfredi C, Ceci A. Innovative study design for paediatric clinical trials. European Journal of Clinical Pharmacology 2011;67:109–15.

[10] Greenberg RG, Gamel B, Bloom D, Bradley J, Jafri HS, Hinton D, et al. Parents' perceived obstacles to pediatric clinical trial participation: Findings from the clinical trials transformation initiative. Contemporary Clinical Trials Communications 2018;9:33–9.

[11] Huff RA, Maca JD, Puri M, Seltzer EW. Enhancing pediatric clinical trial feasibility through the use of Bayesian statistics. Pediatric Research 2017;82:814.

[12] Joseph PD, Craig JC, Caldwell PH. Clinical trials in children. British Journal of Clinical Pharmacology 2015;79:357–69.

[13] Billingham L, Malottki K, Steven N. Small sample sizes in clinical trials: a statistician's perspective. Clinical Investigation 2012;2:655–7.

[14] Kitterman DR, Cheng SK, Dilts DM, Orwoll ES. The prevalence and economic impact of low-enrolling clinical studies at an academic medical center. Academic Medicine: Journal of the Association of American Medical Colleges 2011;86:1360.

[15] ICH E. Clinical investigation of medicinal products in the paediatric population. Note for Guidance on Clinical Investigation of Medicinal Products in the Paediatric Population (CPMP/ICH/2711/99) London 2000.

[16] Use C for MP for H. Guideline on the clinical development of medicinal products for the treatment of cystic fibrosis. London, European Medicines Agency 2009.

[17] Agency EM. Guideline on the requirements for clinical documentation for orally inhaled products (OIP) including the requirements for demonstration of therapeutic equivalence between two inhaled products for use in the treatment of asthma and chronic obstructive pulmonary disease (COPD) in adults and for use in the treatment of asthma in children and adolescents. European Medicines Agency London; 2009.

[18] Products C for PM. Note for Guidance on Evaluation of Anticancer Medicinal Products in Man. The European Agency for the Evaluation of Medicinal Products, London 1996.

[19] O'Hagan A. Bayesian statistics: principles and benefits. Frontis 2004:31–45.

[20] Gajewski BJ, Simon SD, Carlson SE. Predicting accrual in clinical trials with Bayesian posterior predictive distributions. Statistics in Medicine 2008;27:2328–2340.

[21] O'Hagan A. Eliciting expert beliefs in substantial practical applications (Disc: p55-68). Journal of the Royal Statistical Society, Series D: The Statistician 1998;47:21–35.

[22] Lilford RJ, Thornton J, Braunholtz D. Clinical trials and rare diseases: a way out of a conundrum. Bmj 1995;311:1621–5.

[23] Quintana M, Viele K, Lewis RJ. Bayesian Analysis: Using Prior Information to Interpret the Results of Clinical Trials. Jama 2017;318:1605–6.

[24] Comment P. Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials n.d.

[25] Gelman A. Prior distribution. Encyclopedia of Environmetrics 2002.

[26] Albert J. Bayesian computation with R. Springer Science & Business Media; 2009.

[27] Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Bayesian data analysis. vol. 2. CRC press Boca Raton, FL; 2014.

[28] Kawasaki Y, Shimokawa A, Miyaoka E. Comparison of three calculation methods for a Bayesian inference of P ($\pi 1 > \pi 2$). Journal of Modern Applied Statistical Methods 2013;12:15.

[29] Barry J. Doing Bayesian data analysis: A tutorial with R and BUGS. Europe's Journal of Psychology 2011;7:778.

[30] Lunn D, Spiegelhalter D, Thomas A, Best N. The BUGS project: Evolution, critique and future directions. Statistics in Medicine 2009;28:3049–67.

[31] R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2015.

[32] Neelon B, O'Malley AJ. Bayesian analysis using power priors with application to pediatric quality of care. Journal of Biometrics and Biostatistics 2010;1:1–9.

[33] Ibrahim JG, Chen M-H. Power prior distributions for regression models. Statistical Science 2000;15:46–60.

[34] Huang Y-Y, Chen M-J, Chiu N-T, Chou H-H, Lin K-Y, Chiou Y-Y. Adjunctive oral methylprednisolone in pediatric acute pyelonephritis alleviates renal scarring. Pediatrics 2011:peds-2010.

[35] Santis FD. Using historical data for Bayesian sample size determination. Journal of the Royal Statistical Society: Series A (Statistics in Society) 2007;170:95–113. doi:10.1111/j.1467-985X.2006.00438.x.

[36] Bavdekar SB. Pediatric clinical trials. Perspectives in Clinical Research 2013;4:89.

[37] Gill D, Kurz R. Practical and ethical issues in pediatric clinical trials. Applied Clinical Trials 2003;12:41–5.

[38] Carlin BP, Louis TA. Bayesian methods for data analysis. CRC Press; 2008.

[39] Administration UF and D. Guidance for the use of Bayesian statistics in medical device clinical trials. Maryland: US Food and Drug Administration 2010.

Table 1 Credible Interval (95%) for |δ| posterior estimates provided in Informative, Low Informative and Uninformative scenarios.

	Median	Lower	Upper
Power prior without discounting (Informative)	0.26	0.05	**0.46**
Power-prior 50% discounting (Low Informative)	0.26	0.02	**0.49**
Power-prior 100% discounting (Uninformative)	**0.14**	**0.01**	0.44

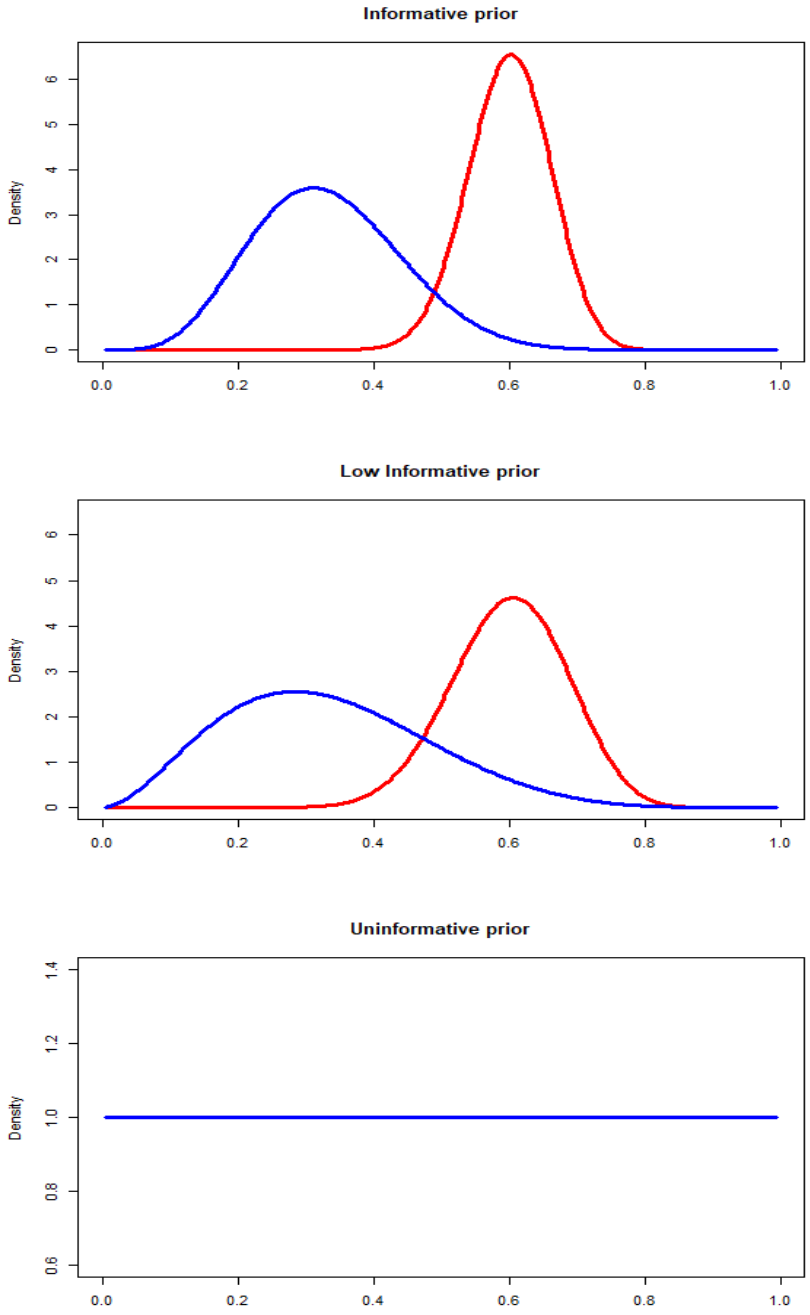

Figure 1 Event rate prior distributions according to treatment in each sensitivity scenario. Colors and arm are not associated, leaving them undisclosed.

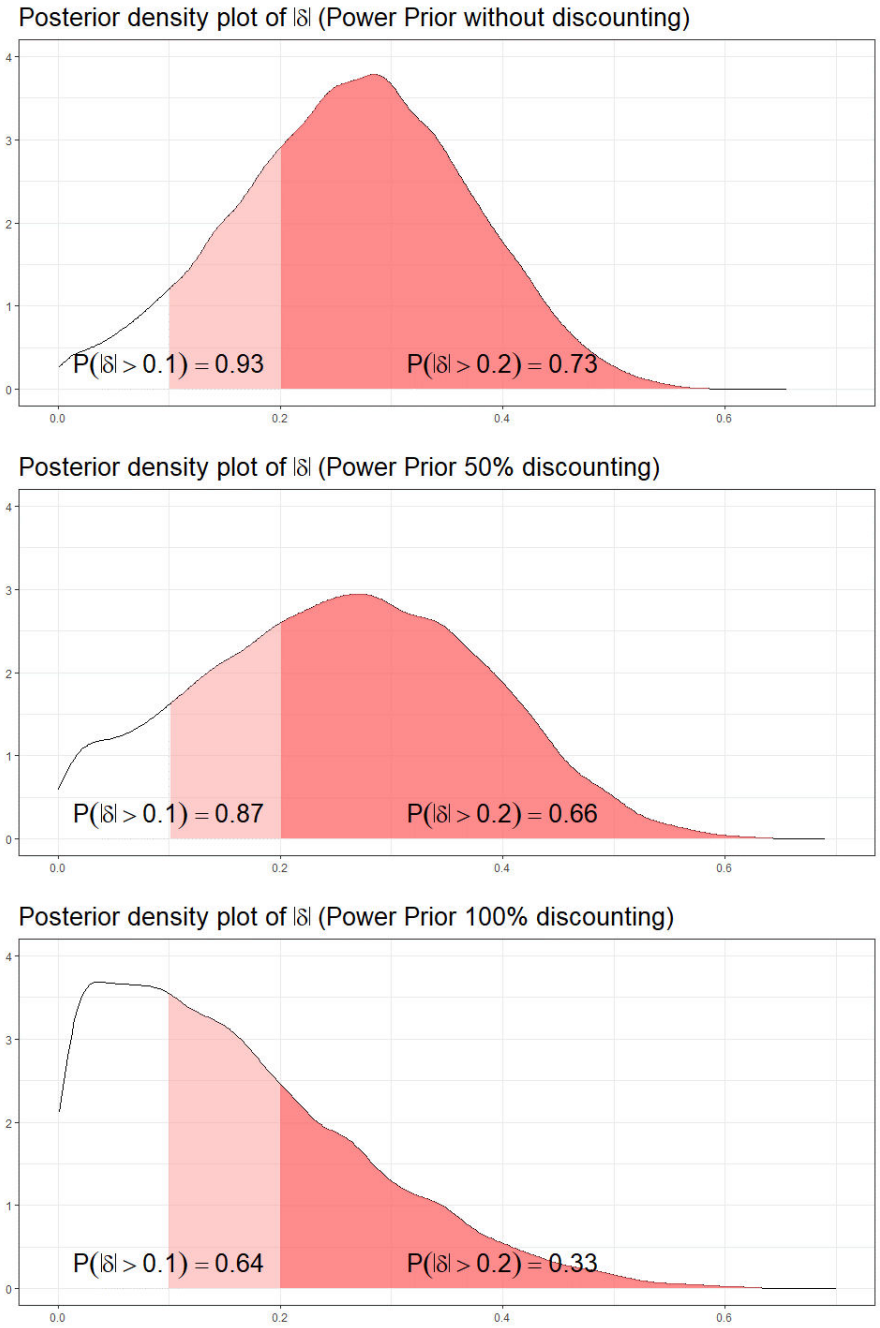

Figure 2 Posterior distributions of |δ| *in each sensitivity scenario.*

Chapter 2

Prior Elicitation for use in clinical trial design and analysis: a literature review[3]

Danila Azzolina[12], Paola Berchialla[3], Ileana Baldi[1]

[1] Unit of Biostatistics, Epidemiology and Public Health, Department of Cardiac Thoracic Vascular Sciences and Public Health, University of Padova, Italy

[2] Department of Statistics, Computing, Applications "Giuseppe Parenti", University of Florence

[3] Department of Clinical and Biological Sciences, University of Torino, Italy

Corresponding author:

Ileana Baldi: Unit of Biostatistics, Epidemiology and Public Health, Department of Cardiac Thoracic Vascular Sciences and Public Health, Via Loredan 18, 35121 Padova, E-mail: ileana.baldi@unipd.it

[3]The article has been submitted to International.Statistical Review

Abstract

The subjective knowledge derived from an expert elicitation may be useful to define a prior distribution when no or limited empirical data are available. This work aims to investigate state-of-the-art Bayesian prior elicitation methods, distinguishing between approaches available in theoretical literature and methods applied in clinical trial research.

A literature search on the Current Index to Statistics (CIS), PubMed and Web of Science (WOS) databases, considering "prior elicitation" search string was run on 5 January 2018.

The paper pertinence was based on the title and the abstract. Summary statistics and time publication trends were reported. Finally, a Latent Dirichlet Allocation (LDA) model was developed to recognize latent topics in the pertinent papers.

Three hundred twenty-five documents pertinent to the Bayesian prior elicitation were identified. Of these, 154 (47%) were published in the "Probability and Statistics" area. Twenty-nine articles pertain to the clinical trial; the majority of them (83%) report the use of parametric techniques for the prior elicitation.

The last decade has seen an increased interest in prior elicitation and the gap between the theory and its application is becoming narrower.

Given the promising flexibility of non-parametric approaches, more efforts are needed to ensure their diffusion in clinical trial settings.

Introduction

The frequentist inference paradigm has been the main statistical approach to the design and analysis of clinical trials since the 1940s [1].

However, the improvements in statistical computing methods and the introduction of the Markov Chain Monte Carlo (MCMC) algorithm have facilitated the spread of Bayesian methods in the field of clinical trials [2].

The prior distribution is a key element of Bayesian inference and represents information about a parameter of interest that is combined with its likelihood to yield the posterior distribution. The prior information may be derived from either expert beliefs (subjective prior) or relevant empirical data (objective prior) [3,4].

Especially when few data are available to estimate the likelihood, for example in clinical trials for rare diseases [5] and poor accrual settings [6], an informative inference complemented with an expert elicitation procedure may be useful to translate the available expert knowledge into a prior probability distribution about treatment effects [7,8]. Clinicians often consult experts to guide clinical practice, particularly when no definitive data are available [9]. Eliciting experts' opinions, in the Bayesian paradigm, may demonstrate the presence of uncertainty in beliefs about treatment effects in a quantifiable and illustrative manner. Moreover, this information can be used to plan a study design, for example, sample size calculations [10] and interim analysis [11]. Elicited prior distributions can be used to augment the information given by scarce therapeutic data [8].

Moreover, it is interesting to consider that the development of user-friendly interfaces, as SHELF (SHeffield ELicitation Framework) [12] or MATCH (Multidisciplinary Assessment of Technology for Healthcare) [13] software, facilitate the application of prior elicitation in both clinical research and in other applied settings.

The SHELF software carries out elicitations of probability distributions for uncertain quantities from a group of experts. Each expert provides a small number of probability opinions corresponding to points on a cumulative distribution function. The SHELF tool fits a range of parametric distributions displaying them in the form of fitted probabilities and percentiles. For multiple experts, a weighted linear pool of the subjective distributions can be calculated [12].

Another useful tool provided in the literature is MATCH, which provides a web-based interface for the SHELF routine but is more user-friendly, including features to remotely conduct the elicitation process [13].

The elicitation process is usually performed by asking experts to report a few summaries of treatment effects; generally, medians, modes, and percentiles of the probability distribution.

Some authors have determined that the role of a facilitator is fundamental in the elicitation process. The facilitator translates the percentiles, defined by experts, into a probability distribution. This process is generally based on parametric distributions (Gamma or Beta, Student, Normal or Log-Normal) [14].

This task is more complicated when opinions are gathered from several experts. In this case, each expert opinion may be separately translated into a distribution. Finally, it is possible to pool each individual distribution into a unique prior distribution.

The elicitation approach accounts for the subjective expert's uncertainty about the treatment effects under investigation, and the consequences of this uncertainty in the final inference can be investigated using sensitivity analysis techniques [8].

Quantiles information about expert beliefs is generally easier to elicit than moments [15]. Probability distributions are, in several cases, defined by moments, and some authors have investigated procedures to derive the parameters of a distribution using the mean and standard deviation [16]. However, instead of considering direct estimates of the mean and standard deviation, it is possible to ask an expert for a specific discrete set of points on the distribution for example quantiles [17]. The mean and standard deviations can be derived by applying specific weights to the quantiles [18], or fitting distributions on to the discrete points [19].

Quantiles information are widely adopted to fit prior probability distributions, not only in the parametric but also in the semiparametric and non-parametric settings; for example, it is possible to ask an expert for the quantiles (usually at least two) of a subjective prior distribution. These points may be plotted, and it is possible to smooth a distribution function drawn through them using a semiparametric or non-parametric representation of the expert's opinion [20,21].

In a parametric setting, the elicitation process assumes that experts' opinions may be represented by a good note family of probability distributions identified by hyper-parameters. Thus, the elicitation consists of the definition of appropriate values for hyper-parameters to represent the experts' beliefs [7].

It is widely assessed that the main limit of a parametric approach is to constrain expert beliefs into a pre-specified distribution [22]. Therefore, non-parametric and semi-parametric hybrid approaches have also been proposed in the elicitation process [21].

This work aims to investigate state-of-the-art of Bayesian prior elicitation methods, focusing on the discrepancy between the available methodological approaches in the statistical literature and the elicitation procedures applied within clinical trial research.

In this general framework, another issue is the identification of the main research topics and the definition of the peculiarities of papers using parametric and non-parametric approaches in a clinical trial with respect to the identified research themes.

Methods

Search strategy

A search on the Current Index to Statistics (CIS), PubMed and Web of Science (WOS) electronic databases, finalized to identify all papers addressing prior elicitation published from 01/01/1980 to 01/05/2018, was performed. The search string “prior elicitation” was used. Papers' pertinence to the inclusion criteria was evaluated based on the title and abstract.

Overall data description

Summary statistics were reported to describe the corpus of papers pertinent to the prior elicitation theme showing the frequency of articles published in journals assigned to the “Probability and Statistics” subject area (here in after referred to as Statistical papers) over time according to the Journal Citation Reports® [23]

As for articles concerning clinical trials, the frequency of published papers has been reported according to publication time and prior elicitation methods, in parametric and in non- (or semi) parametric settings.

State-of-the-art of prior elicitation in clinical trials

Methodological approaches to the prior elicitation techniques currently used in clinical trial literature have been described, evaluating the main characteristics of parametric and non-parametric approaches adopted in trial design and analysis distinguished by type of outcome considered in the study.

For a general comparison purpose, available methods for expert elicitation in the overall pertinent prior elicitation literature have also been reported and described.

Text Mining Analysis

The titles and abstracts of pertinent papers were pre-processed. Punctuation, stop words, white spaces and numbers were removed. Redundant words (e.g., prior, elicitation, expert, or Bayesian, analysis) were also removed. All words were converted to lowercase. Once the text corpus was cleaned, a Document-Term Matrix (DTM) was created. A DTM is a matrix reporting documents (articles) by rows and words by columns; a generic element of DTM is the word counts.

To detect topics, a Latent Dirichlet Allocation (LDA) [24] analysis was performed on the DTM matrix of pertinent articles. The LDA is a technique leading to the automatic discovery of themes in a collection of documents. The method assumes that each document (articles) is a mixture of topics. Documents and words are observed element-wise for topics, which are latent structures discovered by the LDA algorithm.

The method aims to infer the latent topic structure given the words and the document. The LDA recreates the documents in the corpus by adjusting the relative importance of topics in documents iteratively using a Gibbs sampler algorithm [25].

Gibbs sampling operates by performing a random walk; the starting point of the walk is chosen at random; for this reason, it may be useful to discard the first steps (the burn-in period). Overall, 10000 iterations were considered in the computation, and 100 draws were discarded as burn-in. Five Markov chains with different starting points were generated.

The number of topics was chosen following the maximization criterion of the Deveaud measure [26]. The convergence of the LDA algorithm was evaluated showing the Log-Likelihood in correspondence of the first 500 iterations.

Articles were automatically classified according to the relevant topic. The accuracy of the algorithm was evaluated by reading and manually classifying the pertinent trial articles.

Computations were performed using the R 3.3.2 [27] System with topicmodels [28] package.

Results

Overall data description

Twelve hundred and sixty articles were found performing the literature review. Among them, 325 articles are identified as pertinent to the Bayesian prior elicitation theme (Figure 1). Of these, 154 are retrieved in Statistical Journals according to the Journal Citation Reports® classification [23].

As to the temporal pattern of the prior elicitation literature, it is possible to observe that, until 2010, there is a greater number of publications in the statistical literature compared to other research areas; the pattern is reversed from 2010 to January 2018 (**Error! Reference source not found.**).

Concerning the clinical trial research setting, it is possible to observe that 29 articles out of 325 address this research argument. Moreover, according to temporal trends, an increase in the number of publications concerning clinical trials is observed over time; 2 articles between 1991 and 1999, 9 in the period between 2000 and 2009, and 18 between 2009 and 2018.

State-of-the-art prior elicitation in the clinical trial

Table 1 shows the characteristics of the 29 papers pertinent to clinical trial literature.

Parametric approaches

Continuous and time to event outcomes

Considering continuous and time-to-event outcomes, normal or log-normal priors are the preferred distributions for the elicitation procedure in 13 research articles.

The normal distribution is a solution used to define priors on hyper-parameters for a survival function assuming a Weibull time-to-event shaped relation [29]. In several cases, log-hazard ratios are also modelled as a normal distribution [30], for example, in cancer survival studies [31,32]; a multivariate normal distribution has also been used to model the log-hazard of cardiovascular death in studies conducted on a subgroup of patients [33]. This kind of random variable is also considered in the literature to perform a prior elicitation for a survival function in the context of a Bayesian sample size estimation in clinical trial planning [34].

The normal approximation of experts' opinions is also adopted to model parameters of Bayesian logistic regression [35].

In other cases, continuous outcomes, defined on log scales, are modelled eliciting experts' opinions with normal distributions [36]. Additionally, cost data, typically highly skewed, are elicited using log-normal transformations [37].

Multivariate distributions are also used considering an elicitation process based on a mixture of normal distributions [38]. A combination of log-normal priors is also considered to elicitate a prior distribution of internal and external sources of bias affecting final estimates that may be reported to summarize evidence of a randomized clinical trial [39].

Other parametric distributions are considered in the literature for continuous outcomes; for example, surgical learning curve parameters (first procedure and plateau level) are obtained by averaging different experts' opinions using a power law function [40].

Other parametric methods for continuous outcome data are provided in the literature; for example, Inverse Gamma distributions are elicited to model accrual rates in a clinical trial. In another case, Laplace's and Jeffreys's priors are elicited to estimate a competing risks model with covariates [41].

Categorical outcomes

A sequential update of experts' opinions may be reported using, for example, a SHELF elicitation procedure on the event rate [42].

Generally, prior probability distributions for binary outcomes are elicited in term of Beta priors [5,43], but a parametric distribution is also adopted using the log transformation of odds ratios modelling binary data using elicited normal priors [5].

In some cases, a normal distribution has been assumed for parameters characterizing the dose-toxicity curve in a Phase I clinical trial [44]. A phase I clinical trial generally aims to find the maximum tolerated dose, which is often a monotonically increasing dose-response curve following a logistic distribution. For example, the definition of a toxicity response may be based on the approach of eliciting a range for the probability of toxicities at the lowest dose level and the value of the maximum tolerated dose. The prior for both the parameters' distributions may be considered as a Uniform distribution over these ranges [45]. A non-parametric shape function for a maximum tolerated dose may also be reported. Another option addressed in the literature is the elicitation of the toxicity probability at each dose level considering a Beta prior distribution [46].

Not parametric approaches

Five articles [9,47–50] out of the 29 considering expert elicitation in clinical trials consider non-parametric methods for the elicitation of expert opinions.

A graphical visualization of the experts' opinions in histogram form, defined by parameters of a log-hazard function, is a possible approach used to perform elicitation of the expert opinion. The method is flexible, leading to defined hazard regression coefficients with parametric distributions and allowing for non-parametric adjustments using more general copula combinations of marginal distributions [47].

Individual expert histograms representing the prior beliefs about the treatment effects are also used in other cases to derive non-parametric prior averaging individual expert opinions [9,48].

Non-parametric approaches are also used to find the maximum tolerated dose in Phase I clinical trials using the Continual Reassessment Method design, and proposing a suitable informative prior distribution on the relationship between outcome data and covariates [49,51]. In a dose-finding trial, non-parametric elicitation procedures are used to elicit expert quantiles opinions corresponding to the toxicity probability at each dose level [50].

Recently, some efforts are observed in the literature to incorporate alternative procedures to the prior definition in the study design phase. The method is tailored on a phase IIA trial and represents the Bayesian counterparts of a Simon two-stage design using historical data and semi-parametric prior's elicitation methods [52].

State-of-the-art prior elicitation in overall pertinent literature

Concerning generic prior elicitation pertinent literature, several non-parametric methods are provided. For the elicitation of dose-response relations, more flexible approaches are available based on the non-parametric estimation of continuous monotone functions [53].

Also considering the time-to-event endpoints, a semiparametric model is defined on the hazard function as a realization of a stochastic process. The parametric component includes a regression or heterogeneity parameter that is a prior distribution having unknown hyper-parameters (generally, a Gamma or Dirichlet prior process) [54]. A non-parametric prior elicitation method is reported in the literature as an extension of the procedure previously indicated using a tuning factor controlling the degree of the parametric nature of the hazard function [55].

Parameters involved in a logistic regression model may be elicited using a more flexible non-parametric distribution allowing for the possibility that any number of distributions could fit the same judgements [56].

Generally, other methods for continuous outcomes are reported to take into account uncertainty in the prior elicitation procedure. Among them, a method based on using the roulette method for eliciting an expert's probability by providing probabilities of the uncertain quantity of interest [12].

Other methods relied on semiparametric expert elicitation based on B-Splines on expert quantiles; this method is also applied to binary data, but it may also be adapted to other kinds of outcomes [21].

Topic model analysis

The analysis was performed on textual data of 325 articles. The maximum value of the Deveaud metric is 2.19 and has been reached in correspondence of two topics.

The stabilization of the Log-Likelihood measure was reached after the first 100 iterations, indicating an early convergence of the Gibbs sampler algorithm (Figure 3).

The features pertinent to each topic are shown in

Table *2*. Observing the most pertinent word on each topic, we may speculate that:

1. The first topic is more related to the theoretical implications of the prior elicitation procedure (here in after referred to as Theoretical topic).
2. The second topic seems to be related to the empirical application of the prior elicitation methods (here in after referred to as Applied topic).

Table 1 shows that 18 papers are manually classified as applied works (Applied topic), and 11 papers concern a Theoretical topic. All the papers that adopt non-parametric methods are correctly classified by the LDA algorithm; two papers are Applied works, and the other two papers are Theoretical papers. The overall accuracy computed on the manually screened 29 trial articles is equal to 83% (5 articles were misclassified by the LDA algorithm).

Observing the predictions of the LDA algorithm according to publication year (Figure 4), it is possible to observe that the prior elicitation procedure is prevalently addressed in Theoretical topic literature until 2010. The pattern is reversed in recent years, demonstrating an increasing interest in prior elicitations methods in the generally applied research literature.

Moreover, comparing the LDA results about trial articles with the overall pertinent literature on prior elicitation, it is possible to observe a greater proportion of applied papers in trial pertinent literature, and there is evidence that a consistent part of Theoretical literature is allocated in irrelevant articles (Table 3).

Discussion

Study findings indicate that, starting from 2010, it is possible to observe a diffusion of the prior elicitation techniques into research fields different from theoretical statistics. This finding may be related to the recent increase in popularity of Bayesian methods in a general setting and in clinical trial research [2]. In recent years, Bayesian methods have increasingly been used in the design, monitoring, and analysis of clinical trials due to their flexibility [57].

The increase in popularity of Bayesian methods in clinical trials involves a need for statisticians to define the tools that are useful for the definition of robust and defensible informative prior distributions [58].

Empirical data may be used to define such priors (objective prior) whenever possible. However, in some cases, the limitations in data availability may preclude the construction of a data-based prior. In this situation, an expert elicitation procedure may be a solution used to define prior distributions [58].

In clinical trial publications, most of the literature is related to the field of theoretical statistics. This aspect especially concerns less used approaches involving non-parametric methods for prior elicitation methods. The reason behind the limited application of the non-parametric methods is clearly related to the computational effort associated with the definition of a prior distribution, which is more flexible and adaptable to the expert opinion, but, in several cases, leads to obtaining posterior distributions that are difficult to express in the closed form [59].

It is important to consider that, in some research contexts, the translation of the experts' opinions into a pre-specified family distribution may be considered a limitation because many different distributions may be more suitable to the experts' opinions that are generally expressed in quantiles [14].

In recent years, not only parametric but also non-parametric methods for the elicitation of expert opinion have been examined, especially in the theoretical literature.

However, in clinical trial research, the conventional parametric methods are the more adopted procedures for the elicitation of expert opinions, leaving non-parametric methods predominantly in statistical fields.

Given the potential of prior elicitation for better decision making, more efforts are needed to ensure the diffusion of prior elicitation facilities, not only in theoretical statistical research but also in applied clinical trial settings, both at the design and analysis stage.

Conclusion

Prior elicitation methods are recently appealing not only to the statistical literature but also in other research settings. It is possible to observe that the methods are increasingly being used in the general literature and clinical trial research.

However, in this framework, conventional parametric methods are more popular in clinical trial research. Non-parametric approaches are, in several cases, treated specially in the theoretical literature, which is mainly focused on statistical argumentations.

References

[1] Lee JJ, Chu CT. Bayesian clinical trials in action. Stat Med 2012;31:2955–72. doi:10.1002/sim.5404.

[2] Chevret S. Bayesian adaptive clinical trials: a dream for statisticians only? Statistics in Medicine 2012;31:1002–13.

[3] Chaloner K, Rhame FS. Quantifying and documenting prior beliefs in clinical trials. Statistics in Medicine 2001;20:581–600.

[4] Dolan JG, Bordley DR, Mushlin AI. An Eualuation of Clinicians' Subjective Prior Probability Estimates. Medical Decision Making 1986;6:216–23.

[5] Hampson LV, Whitehead J, Eleftheriou D, Brogan P. Bayesian methods for the design and interpretation of clinical trials in very rare diseases. Statistics in Medicine 2014;33:4186–201. doi:10.1002/sim.6225.

[6] Quintana M, Viele K, Lewis RJ. Bayesian Analysis: Using Prior Information to Interpret the Results of Clinical Trials. Jama 2017;318:1605–6.

[7] Garthwaite PH, Kadane JB, O'Hagan A. Statistical methods for eliciting probability distributions. Journal of the American Statistical Association 2005;100:680–701.

[8] O'Hagan A. Eliciting expert beliefs in substantial practical applications (Disc: p55-68). Journal of the Royal Statistical Society, Series D: The Statistician 1998;47:21–35.

[9] Johnson SR, Granton JT, Tomlinson GA, Grosbein HA, Hawker GA, Feldman BM. Effect of warfarin on survival in scleroderma-associated pulmonary arterial hypertension (SSc-

PAH) and idiopathic PAH. Belief elicitation for Bayesian priors. The Journal of Rheumatology 2011;38:462–9. doi:10.3899/jrheum.100632.

[10] Spiegelhalter D, Freedman LS. A predictive approach to selecting the size of a clinical trial, based on subjective clinical opinion. Statistics in Medicine 1986;5:1–13.

[11] Spiegelhalter D. Incorporating Bayesian ideas into health-care evaluation. Statistical Science 2004;19:156–74.

[12] Gosling JP. SHELF: The Sheffield Elicitation Framework. In: Dias LC, Morton A, Quigley J, editors. Elicitation, vol. 261, Cham: Springer International Publishing; 2018, p. 61–93. doi:10.1007/978-3-319-65052-4_4.

[13] Morris DE, Oakley JE, Crowe JA. A web-based tool for eliciting probability distributions from experts. Environmental Modelling & Software 2014;52:1–4.

[14] O'Hagan A, Buck CE, Daneshkhah A, Eiser JR, Garthwaite PH, Jenkinson DJ, et al. Uncertain judgements: eliciting experts' probabilities. Chichester: John Wiley. ISBN; 2006.

[15] Kiefer NM. Incentive-compatible elicitation of quantiles. ArXiv Preprint ArXiv:161100868 2016.

[16] Lau H-S, Lau AH-L. An improved PERT-type formula for standard deviation. IIE Transactions 1998;30:273–5.

[17] Zapata-Vazquez RE, O'Hagan A, Bastos LS. Eliciting expert judgements about a set of proportions. Journal of Applied Statistics 2014;41:1919–33. doi:10.1080/02664763.2014.898131.

[18] Keefer DL. Certainty equivalents for three-point discrete-distribution approximations. Management Science 1994;40:760–73.

[19] Abbas AE, Budescu DV, Yu H-T, Haggerty R. A comparison of two probability encoding methods: Fixed probability vs. fixed variable values. Decision Analysis 2008;5:190–202.

[20] Winkler RL. The assessment of prior distributions in Bayesian analysis. Journal of the American Statistical Association 1967;62:776–800.

[21] Bornkamp B, Ickstadt K. A note on B-splines for semiparametric elicitation. The American Statistician 2009;63:373–7.

[22] Johnson SR, Tomlinson GA, Hawker GA, Granton JT, Grosbein HA, Feldman BM. A valid and reliable belief elicitation method for Bayesian priors. Journal of Clinical Epidemiology 2010;63:370–83. doi:10.1016/j.jclinepi.2009.08.005.

[23] Reuters T. Journal citation reports. Thomson Reuters; 2011.

[24] Blei DM, Lafferty JD. Topic models. Text Mining, Chapman and Hall/CRC; 2009, p. 101–24.

[25] Porteous I, Newman D, Ihler A, Asuncion A, Smyth P, Welling M. Fast collapsed gibbs sampling for latent dirichlet allocation. Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM; 2008, p. 569–77.

[26] Deveaud R, SanJuan E, Bellot P. Accurate and effective latent concept modeling for ad hoc information retrieval. Document Numérique 2014;17:61–84.

[27] R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2015.

[28] Hornik K, Grün B. topicmodels: An R package for fitting topic models. Journal of Statistical Software 2011;40:1–30.

[29] Chen MH, Ibrahim JG, Sinha D. A new Bayesian model for survival data with a surviving fraction. Journal of the American Statistical Association 1999;94:909–19. doi:10.2307/2670006.

[30] Tan SB, Chung YF, Tai BC, Cheung YB, Machin D. Elicitation of prior distributions for a phase III randomized controlled trial of adjuvant therapy with surgery for hepatocellular carcinoma. Controlled Clinical Trials 2003;24:110–21.

[31] Hiance A, Chevret S, Levy V. A practical approach for eliciting expert prior beliefs about cancer survival in phase III randomized trial. Journal of Clinical Epidemiology 2009;62:431–7. doi:10.1016/j.jclinepi.2008.04.009.

[32] Moatti M, Zohar S, Facon T, Moreau P, Mary JY, Chevret S. Modeling of experts' divergent prior beliefs for a sequential phase III clinical trial. Clinical Trials 2013;10:505–14. doi:10.1177/1740774513493528.

[33] White IR, Pocock SJ, Wang D. Eliciting and using expert opinions about influence of patient characteristics on treatment effects: a Bayesian analysis of the CHARM trials. Statistics in Medicine 2005;24:3805–21. doi:10.1002/sim.2420.

[34] Ren S, Oakley JE. Assurance calculations for planning clinical trials with time-to-event outcomes. Statistics in Medicine 2014;33:31–45. doi:10.1002/sim.5916.

[35] O'Leary RA, Choy SL, Murray JV, Kynn M, Denham R, Martin TG, et al. Comparison of three expert elicitation methods for logistic regression on predicting the presence of the threatened brush-tailed rock-wallaby Petrogale penicillata. Environmetrics 2009;20:379–98. doi:10.1002/env.935.

[36] See CW, Srinivasan M, Saravanan S, Oldenburg CE, Esterberg EJ, Ray KJ, et al. Prior Elicitation and Bayesian Analysis of the Steroids for Corneal Ulcers Trial. Ophthalmic Epidemiology 2012;19:407–13. doi:10.3109/09286586.2012.735332.

[37] Stevens JW, O'Hagan A. Incorporation of genuine prior information in cost-effectiveness analysis of clinical trial data. International Journal of Technology Assessment in Health Care 2002;18:782–90.

[38] Thall PF, Ursino M, Baudouin V, Alberti C, Zohar S. Bayesian treatment comparison using parametric mixture priors computed from elicited histograms. Statistical Methods in Medical Research 2017:962280217726803. doi:10.1177/0962280217726803.

[39] Turner RM, Spiegelhalter DJ, Smith GCS, Thompson SG. Bias modelling in evidence synthesis. Journal of the Royal Statistical Society Series A-Statistics in Society 2009;172:21–47. doi:10.1111/j.1467-985X.2008.00547.x.

[40] Cook JA, Ramsay CR, Carr AJ, Rees JL. A questionnaire elicitation of surgeons' belief about learning within a surgical trial. PloS One 2012;7:e49178. doi:10.1371/journal.pone.0049178.

[41] Coolen FPA, Mertens PR, Newby MJ. A Bayes-competing risk model for the use of expert judgment in reliability estimation. Reliability Engineering and System Safety 1992;35:23–30.

[42] Higgins HM, Dryden IL, Green MJ. A Bayesian approach demonstrating that incorporation of practitioners' clinical beliefs into research design is crucial for effective knowledge transfer. Udder Health and Communication, Springer; 2011, p. 133–40.

[43] Rovers MM, van der Wilt GJ, van der Bij S, Straatman H, Ingels K, Zielhuis GA. Bayes' theorem: A negative example of a RCT on grommets in children with glue ear. European Journal of Epidemiology 2005;20:23–8. doi:10.1007/s10654-004-1594-y.

[44] Bekele BN, Thall PF. Dose-finding based on multiple toxicities in a soft tissue sarcoma trial. Journal of the American Statistical Association 2004;99:26–35. doi:10.1198/016214504000000043.

[45] Rosenberger WF, Canfield GC, Perevozskaya I, Haines LM, Hausner P. Development of interactive software for Bayesian optimal phase 1 clinical trial design. Drug Information Journal 2005;39:89–98. doi:10.1177/009286150503900112.

[46] Cheung YK. On the use of nonparametric curves in phase I trials with low toxicity tolerance. Biometrics 2002;58:237–40. doi:10.1111/j.0006-341X.2002.00237.x.

[47] Chaloner K, Church T, Louis TA, Matts JP. Graphical elicitation of a prior distribution for a clinical trial. Journal of the Royal Statistical Society, Series D: The Statistician 1993;42:341–53.

[48] Sun CQ, Prajna NV, Krishnan T, Mascarenhas J, Rajaraman R, Srinivasan M, et al. Expert Prior Elicitation and Bayesian Analysis of the Mycotic Ulcer Treatment Trial I. Investigative Ophthalmology & Visual Science 2013;54:4167–73. doi:10.1167/iovs.13-11716.

[49] Legedza ATR, Ibrahim JG. Heterogeneity in phase I clinical trials: prior elicitation and computation using the continual reassessment method. Statistics in Medicine 2001;20:867–82. doi:10.1002/sim.701.

[50] Zohar S, Baldi I, Forni G, Merletti F, Masucci G, Gregori D. Planning a Bayesian early-phase phase I/II study for human vaccines in HER2 carcinomas. Pharmaceutical Statistics 2011;10:218–26. doi:10.1002/pst.450.

[51] Ibrahim JG, Ryan LM, Chen M-H. Using historical controls to adjust for covariates in trend tests for binary data. Journal of the American Statistical Association 1998;93:1282–93.

[52] Berchialla P, Zohar S, Baldi I. Bayesian sample size determination for phase IIA clinical trials using historical data and semi-parametric prior's elicitation. Pharmaceutical Statistics 2018.

[53] Bornkamp B, Ickstadt K. Bayesian nonparametric estimation of continuous monotone functions with applications to dose–response analysis. Biometrics 2009;65:198–205.

[54] Sinha D, Dey DK. Semiparametric Bayesian analysis of survival data. Journal of the American Statistical Association 1997;92:1195–212.

[55] Sinha D, Chen M-H, Ibrahim JG. Bayesian inference for survival data with a surviving fraction. Lecture Notes-Monograph Series 2003:117–38.

[56] Gosling JP, Oakley JE, O'Hagan A. Nonparametric elicitation for heavy-tailed prior distributions. Bayesian Analysis 2007;2:693–718.

[57] Baldi I, Gregori D, Desideri A, Berchialla P. Accrual monitoring in cardiovascular trials. Open Heart 2017;4:e000720. doi:10.1136/openhrt-2017-000720.

[58] Dallow N, Best N, Montague TH. Better decision making in drug development through adoption of formal prior elicitation. Pharmaceutical Statistics 2018;17:301–16.

[59] Ghahramani Z. Bayesian non-parametrics and the probabilistic approach to modelling. Phil Trans R Soc A 2013;371:20110553.

[60] Browne EN, Rathinam SR, Kanakath A, Thundikandy R, Babu M, Lietman TM, et al. A Bayesian Analysis of a Randomized Clinical Trial Comparing Antimetabolite Therapies for Non-Infectious Uveitis. Ophthalmic Epidemiology 2017;24:63–70. doi:10.1080/09286586.2016.1255764.

[61] Gajewski BJ, Simon SD, Carlson SE. Predicting accrual in clinical trials with Bayesian posterior predictive distributions. Statistics in Medicine 2008;27:2328–40. doi:10.1002/sim.3128.

[62] Hampson LV, Whitehead J, Eleftheriou D, Tudur-Smith C, Jones R, Jayne D, et al. Elicitation of Expert Prior Opinion: Application to the MYPAN Trial in Childhood Polyarteritis Nodosa. Plos One 2015;10. doi:10.1371/journal.pone.0120981.

[63] Higgins HM, Dryden IL, Green MJ. A Bayesian elicitation of veterinary beliefs regarding systemic dry cow therapy: Variation and importance for clinical trial design. Preventive Veterinary Medicine 2012;106:87–96. doi:10.1016/j.prevetmed.2012.01.017.

[64] Rietbergen C, Klugkist I, Janssen KJ, Moons KG, Hoijtink HJ. Incorporation of historical data in the analysis of randomized therapeutic trials. Contemporary Clinical Trials 2011;32:848–55. doi:10.1016/j.cct.2011.06.002.

[65] Veen D, Stoel D, Zondervan-Zwijnenburg M, van de Schoot R. Proposal for a Five-Step Method to Elicit Expert Judgment. Frontiers in Psychology 2017;8. doi:10.3389/fpsyg.2017.02110.

[66] Wang CP, Ghosh M. Bayesian analysis of bivariate competing risks models with covariates. Journal of Statistical Planning and Inference 2003;115:441–59. doi:10.1016/s0378-3758(02)00177-5.

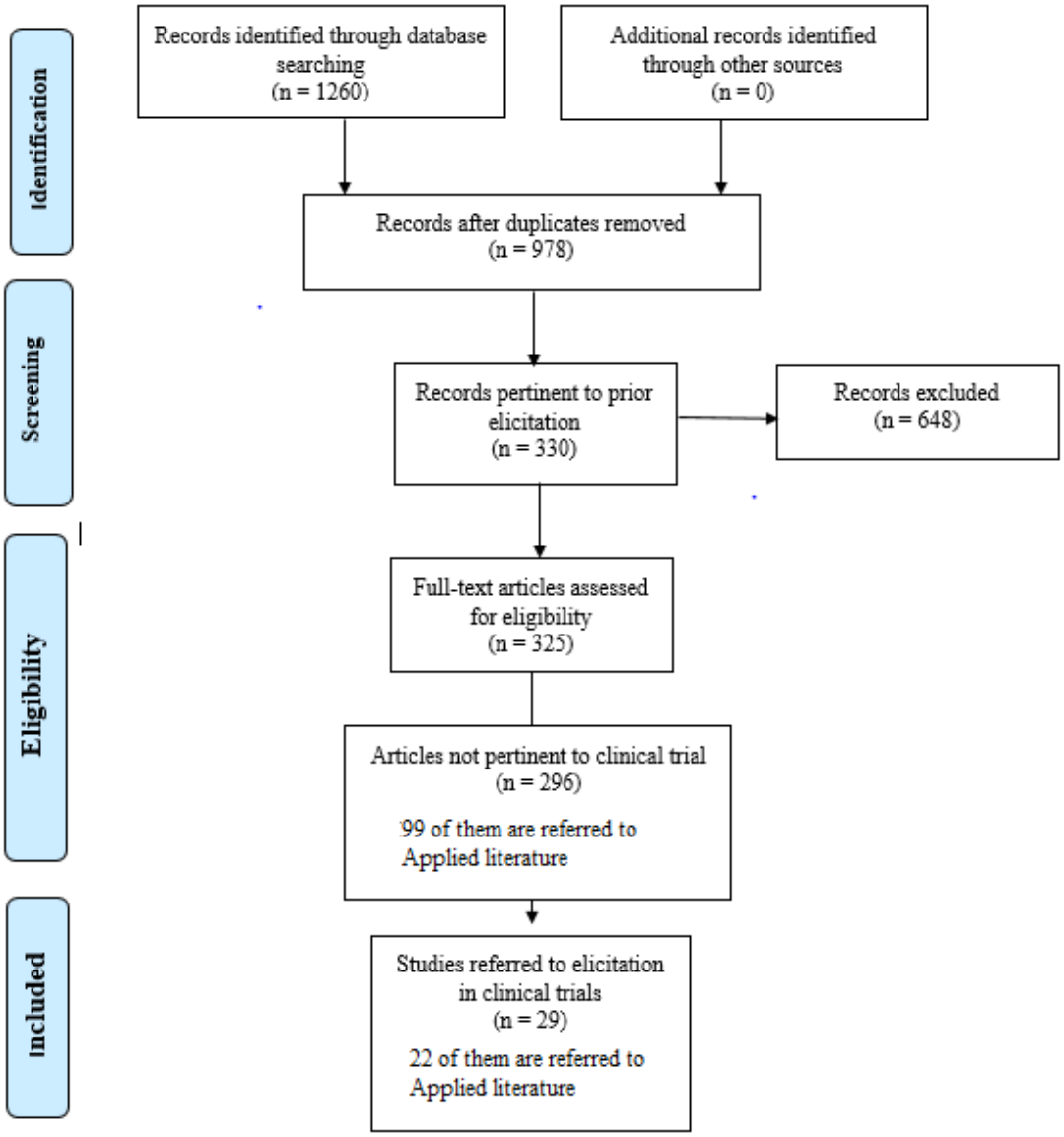

Figure 1 Prisma Flowchart

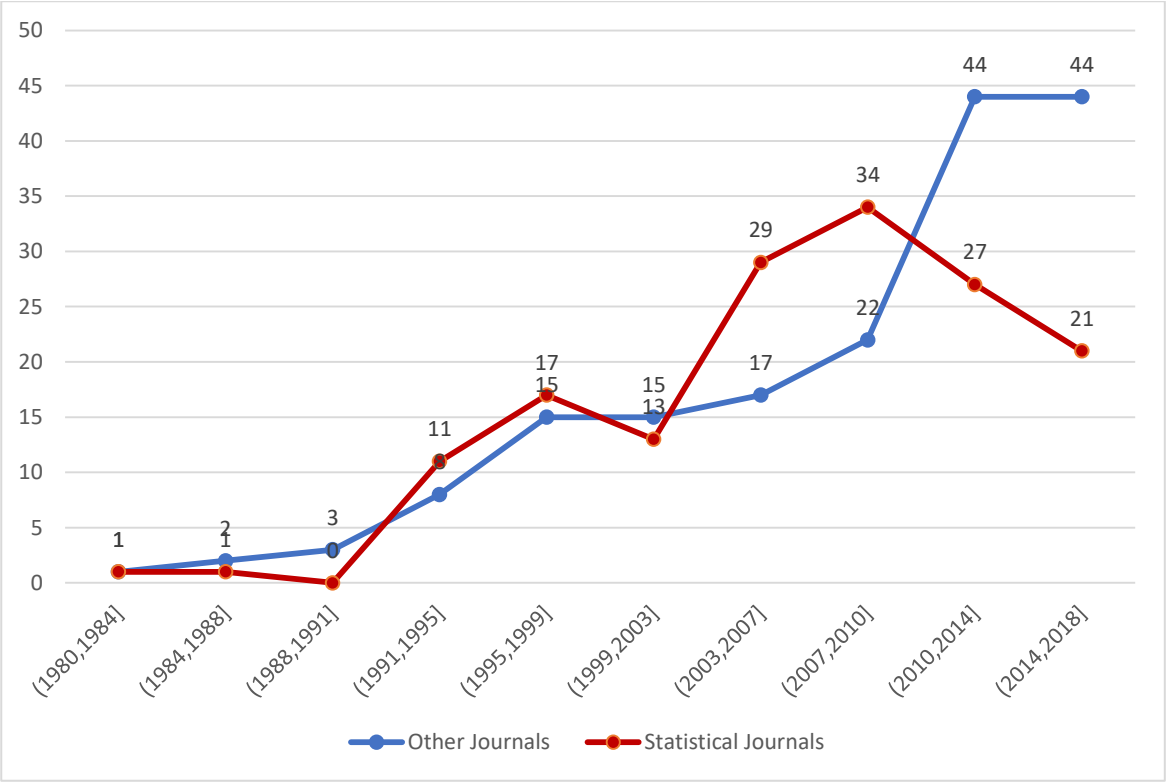

Figure 2 Articles pertinent to prior elicitation (n=325) according to journal type and year

Table 1 Articles regarding prior elicitation in clinical trials classified according to publication year, first author, title, main approach, and prior distribution

Publication Year	Author	Title	Approach	Prior Distribution	Manual Classification
2004	[44]	Dose-finding based on multiple toxicities in a soft tissue sarcoma trial	Parametric	Multinormal	Applied
2017	[60]	A Bayesian Analysis of a Randomized Clinical Trial Comparing Antimetabolite Therapies for Non-Infectious Uveitis	Parametric	Normal	Applied
1993	[47]	Graphical elicitation of a prior distribution for a clinical trial	Not Parametric	Non-parametric adjustments via copula combinations of marginal distribution	Theoretical
1999	[29]	A new Bayesian model for survival data with a surviving fraction	Parametric	Gamma and Normal	Theoretical
2002	[46]	On the use of nonparametric curves in phase I trials with low toxicity tolerance	Parametric	Beta	Theoretical
2012	[40]	A questionnaire elicitation of surgeons' belief about learning within a surgical trial	Parametric	Learning curve	Applied
2008	[61]	Predicting accrual in clinical trials with Bayesian posterior predictive distributions	Parametric	Inverse Gamma	Theoretical
2015	[62]	Elicitation of Expert Prior Opinion: Application to the MYPAN Trial in Childhood Polyarteritis Nodosa	Parametric	Beta and Normal	Applied
2009	[31]	A practical approach for eliciting expert prior beliefs about cancer survival in phase III randomized	Parametric	Normal	Applied

		trial			
2011	[42]	A Bayesian approach demonstrating that the incorporation of practitioners' clinical beliefs into research design is crucial for effective knowledge transfer	Parametric	Beta	Applied
2012	[63]	A Bayesian elicitation of veterinary beliefs regarding systemic dry cow therapy: Variation and importance for the clinical trial design	Parametric	Beta	Applied
2011	[9]	Effect of warfarin on survival in scleroderma-associated pulmonary arterial hypertension (SSc-PAH) and idiopathic PAH. Belief elicitation for Bayesian priors	Not Parametric	Density histogram	Applied
2001	[49]	Heterogeneity in phase I clinical trials: prior elicitation and computation using the continual reassessment method	Not Parametric	Ibrahim Prior [51]	Theoretical
2013	[32]	Modeling of experts' divergent prior beliefs for a sequential phase III clinical trial	Parametric	Mixture of normal	Theoretical
2009	[35]	Comparison of three expert elicitation methods for logistic regression on predicting the presence of the threatened brush-tailed rock-wallaby Petrogale penicillata	Parametric	Normal and Multinormal	Theoretical
2014	[34]	Assurance calculations for planning clinical trials with time-to-event outcomes	Parametric	Log-Normal	Theoretical
2011	[64]	Incorporation of historical data in the analysis of randomized therapeutic trials	Parametric	Beta	Theoretical

2005	[45]	Development of interactive software for Bayesian optimal phase 1 clinical trial design	Parametric	Uniform	Applied
2005	[43]	Bayes' theorem: A negative example of an RCT on grommets in children with glue ear	Parametric	Beta	Applied
2012	[36]	Prior Elicitation and Bayesian Analysis of the Steroids for Corneal Ulcers Trial	Parametric	Mixture of normal	Applied
2002	[37]	Incorporation of genuine prior information in cost-effectiveness analysis of clinical trial data	Parametric	Log-Normal	Applied
2013	[48]	Expert Prior Elicitation and Bayesian Analysis of the Mycotic Ulcer Treatment Trial I	Not Parametric	Density histogram	Applied
2003	[30]	Elicitation of prior distributions for a phase III randomized controlled trial of adjuvant therapy with surgery for hepatocellular carcinoma	Parametric	Normal	Applied
2017	[38]	Bayesian treatment comparison using parametric mixture priors computed from elicited histograms	Parametric	Mixture of Normal	Theoretical
2009	[39]	Bias modeling in evidence synthesis	Parametric	Log-Normal	Applied
2017	[65]	Proposal for a Five-Step Method to Elicit Expert Judgment	Parametric	Normal	Applied
2003	[66]	Bayesian analysis of bivariate competing risks models with covariates	Parametric	Laplace and Jeffreys Prior	Theoretical
2005	[33]	Eliciting and using expert opinions about the influence of patient characteristics on treatment effects: a Bayesian analysis of the	Parametric	Normal	Applied

		CHARM trials			
2011	[50]	Planning a Bayesian early-phase phase I/II study for human vaccines in HER2 carcinomas	Not Parametric	Not parametric expert quantiles distribution	Applied

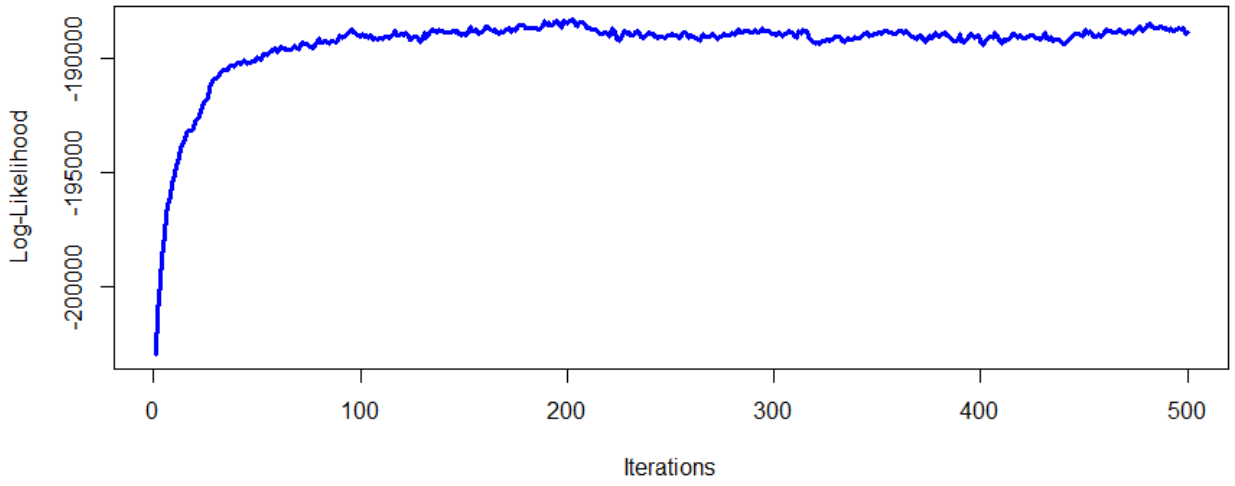

Figure 3 Log-Likelihood reached in correspondence of first 500 iterations of the LDA Gibbs algorithm

Table 2 Pertinent words according to each LDA topic. The most important words are represented in bold

	Theoretical	**Applied**
1	**model**	**studi**
2	**distribut**	**probabl**
3	**data**	**result**
4	**inform**	**estim**
5	paramet	knowledg
6	propos	approach
7	posterior	provid
8	base	assess
9	function	opinion
10	statist	test

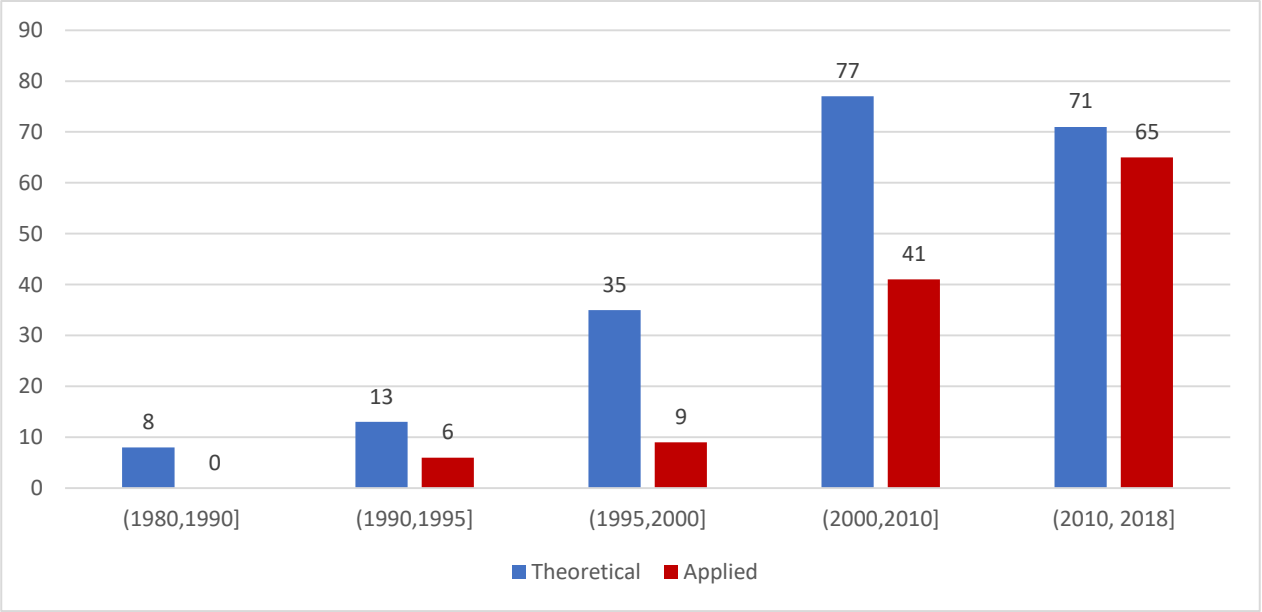

Figure 4 Classification of articles pertinent to prior elicitation according to the LDA topics and publication years.

Table 3 Classification of articles according to LDA topics and pertinence to the clinical trial literature.

	Applied	**Theoretical**	**Total**
Pertinent to clinical trials	18% (22)	3% (7)	29
Not pertinent to clinical trials	82% (99)	97% (197)	296
Total	**121**	**204**	**325**

Chapter 3

A Bayesian sample size estimation procedure based on a B-splines semiparametric elicitation method[4]

Danila Azzolina[1,2], Paola Berchialla[3], Dario Gregori[1], Ileana Baldi[1]

[1] Unit of Biostatistics, Epidemiology and Public Health, Department of Cardiac Thoracic Vascular Sciences and Public Health, University of Padova, Italy

[2] Department of Statistics, Computing, Applications "Giuseppe Parenti", University of Florence

[3] Department of Clinical and Biological Sciences, University of Torino, Italy

Corresponding author:

Ileana Baldi: Unit of Biostatistics, Epidemiology and Public Health, Department of Cardiac Thoracic Vascular Sciences and Public Health, Via Loredan 18, 35121 Padova, E-mail: ileana.baldi@unipd.it

[4] The article will be submitted to Statistics for Biopharmaceutical Research.

Abstract

Sample size estimation is a fundamental element of a clinical trial, and a binomial experiment is the most common situation faced in clinical trial design.

A Bayesian method to determine sample size is an alternative solution to a frequentist design in cases where there is difficulty in the enrollment process and for studies conducted on small sample sizes. The Bayesian method uses available knowledge, which is translated into a prior distribution, and the binomial parameter under investigation, instead of a point estimate. This approach takes the uncertainty in data prediction completely into account.

In the case that there are no objective data available to define a prior distribution, it may be indispensable to use an expert opinion to derive the prior distribution in performing an elicitation procedure. Expert elicitation refers to the process of translating the expert opinion into a prior probability distribution.

We investigate the estimation of a binomial sample size providing a generalized version of the average length and average coverage criteria, and worst outcome criterion. The original method was proposed by Joseph and is defined in a parametric framework based on a beta-binomial model.

In this theoretical setting, we propose a more flexible approach for binary data sample size estimation considering in the method not only parametric approaches (beta priors) but also semiparametric priors based on B-splines.

Introduction

Bayesian trials are increasingly popular in clinical research especially when studies are conducted in poor accrual settings or for studies conducted on rare diseases [1,2]. However, according to international guidelines [3], a Bayesian trial has to be planned from the study design phase, providing an appropriate sample size estimation procedure, before analyzing the data.

In a frequentist framework, clinical trials are generally designed to maximize the probability of correctly rejecting the null hypothesis under evaluation. However, this approach may involve a loss of statistical power in the analysis conducted on smaller sample sizes with respect to indications provided by the study protocol [2].

A Bayesian design, instead, may be useful in this research situation. The method uses the available information about treatment effects, translated into informative prior distributions, in order to reduce the uncertainty effect in size estimation instead of providing a definitive answer to a statistical hypothesis defined in a frequentist study design [2].

Different Bayesian methods are available in the literature to obtain a sample size estimation procedure for binary data based on optimization of the precision, defined on a generic posterior interval or on the posterior variance. Some authors identify the sample size defining a tolerance area R in which the parameters of a Binomial or Multinomial distribution will be contained with a specified probability [4]. Pham-Gia and Turkkan obtained sample sizes for a binomial distribution, in closed form considering a beta-binomial model, by imposing precision conditions on the posterior variance and on the Bayes risk factor [5].

Other precision approaches to sample size estimation are based on optimization of the length and coverage of the HPD (Highest Posterior Density) interval. This kind of posterior interval has the property that any point within the interval has a higher density than any other point outside interval including most likely values of the parameters. However, HPD does not always result in an interval estimate when the posterior density is multimodal, then HPD can yield non-interval set estimators, in contrast to quantile credibility interval [7]. Widely adopted procedures, among HPD sample size estimation methods, are the Average Coverage Criterion (ACC) [8], the Average Length Criterion (ALC) [8], and the Worst Outcome Criterion (WOC) [8].

The ACC fixes the length of the posterior intervals and controls the coverage probability level over the data. The ALC method instead fixes the coverage rate and optimizes the length of intervals among the data space. The WOC, a more conservative criterion, controls both length and coverage of the intervals over all possible data [8]. Other generalizations of the HPD interval sample size

approaches provided in the literature consider the median of coverage and length of the intervals among the data space, or performs WOC computations on a specific subset of data [9].

All the proposed sample size solutions are developed considering a parametric definition of the prior distribution, specifically in a beta-binomial framework for binary data.

The definition of an appropriate prior distribution plays a central role in Bayesian trial design and analysis [10]. Objective data, retrieved by other studies, may be considered to derive informative distributions (objective prior). However, in some cases, the empirical evidence on treatment effect is not available; in this research setting, expert opinions may be translated into an informative prior (elicitation process) [11].

Different methods may be considered to conduct an elicitation procedure in a parametric, semiparametric and nonparametric setting. The parametric methods force the expert's opinion into a prespecified density function characterized by hyper-parameters [12].

Nonparametric elicitation does not make any assumption about the distribution form of the expert opinion, and semiparametric approaches are hybrid solutions [13]. The literature indicates that, in several cases, nonparametric or semiparametric approaches are more pliable methods to the elicitation of expert beliefs [14].

The B-splines semiparametric method, for example, is a very flexible procedure which leads to obtaining a prior distribution by performing a balanced optimization of a weighted sum of two components; one is a linear combination of B-splines adapted among expert's quantiles, another one is an uninformative uniform prior distribution [13].

Recently, some efforts are evidenced in the literature to incorporate alternative procedures to the prior definition during the study design phase. The method is tailored on a phase IIA trial and represents a Bayesian counterpart of a Simon two-stage design using historical data and semi-parametric prior elicitation methods [15].

Instead, this work proposes a generalized version of the ALC, ACC, and WOC, including not only parametric approaches (beta priors) but also semiparametric priors based on B-splines in the sample size estimation method.

The proposed sample size estimation method is also applied to a motivating example, a phase II clinical trial aimed at assessing the effects of pharmacological treatment on a binary safety endpoint in a pediatric population. This kind of study design has generally one sample, a single stage and is conducted on small sample sizes, in which enrolled patients are treated and then observed for a possible response, generally binary [16]. The opinions provided by 8 experts are considered to elicit informative priors used to design the trial in both a parametric beta-binomial and semiparametric B-splines setting.

Methods

Bayesian methods for sample size estimation

Considering an unknown parameter θ, and a parametric space Θ for unknown θ, the prior distribution has a density function $f(\theta)$. The data, considering a sample size equal to *n*, are $x = (x_1, x_2, \dots, x_n)$ and are assumed interchangeable among data space χ.

Different Bayesian sample size criteria are proposed in the literature for binary data:

1. **ACC** [4]**.** The length of the (HPD) interval is fixed, and the coverage level $(1-\alpha)$ varies depending on the sample. The sample size *n* is chosen to obtain average coverage over the data greater or equal to $(1-\alpha)$.
2. **ALC** [8]**.** The main idea is that the coverage probability $(1-\alpha)$ is fixed and the HPD interval length varies according to the sample. The sample size *n* is chosen to be a minimum integer which satisfies the condition that the average width is less than or equal to a prespecified length *l.*
3. **WOC** [8]**.** Is a more conservative approach, ensuring a maximum length *l* and a minimum coverage probability $(1-\alpha)$ of all the data that may occur.

Average Coverage Criterion

Considering an HPD credible interval, it is possible to assume a fixed length *l*. The coverage $(1-\alpha)$ instead varies with the data among the overall data space χ. An ACC sample size is the smallest integer such that, for a length *l*, the expected coverage level is at least $1-\alpha$.

$$\int_{\chi}\left\{\int_{a(x,n)}^{a(x,n)+l} f(\theta|x,n)\mathrm{d}\theta\right\} f(x)\mathrm{d}x \geq 1-\alpha$$

The predictive distribution of the data is the preposterior marginal distribution and is $f(x)$; the posterior distribution is $f(\theta|x)$, defined as combining likelihood information $f(x|\theta)$ and prior distribution $f(\theta)$.

In the equation, *a(x, n)* is the lower bound of the HPD interval of prespecified length *l*, considering a posterior density function $f(\theta|x,n)$ which is related to data *x* and sample size *n*.

The relation reported on the left side may be interpreted as an average of the posterior coverage, weighted by the predictive distribution *f(x)* of the data.

Considering the case of the estimation of the proportion for one sample, the notation for the ACC criteria may be reported as:

$$\sum_{x=0}^{n} Pr\{\theta \in (a(x,n), a(x,n)+l)\}p(x,n) \geqslant 1-\alpha$$

The coverage level may be reported as:

$$Pr\{\theta \in (a(x,n), a(x,n)+l)\} \propto \int_{a(\mathbf{x},n)}^{a(x,n)+l} \theta^{x}(1-\theta)^{(n-x)} f(\theta)\mathrm{d}\theta$$

The relation integrates the product of the Binomial likelihood and the prior $f(\theta)$ over the space comprised between the lower limit $a(\mathrm{x},n)$and the upper limit $a(x,n)+l$ of the posterior or HPD interval.

In the conventional approach, the prior distribution is a beta function where c and d are Shape and Scale parameters.

$$f(\theta) = \frac{1}{B(c,d)}\theta^{(c-1)}(1-\theta)^{(d-1)}, 0<\theta<1$$

In a beta-binomial setting, the posterior is the product between the binomial likelihood and beta prior:

$$f(\theta|x,n,c,d) = \frac{1}{B(x+c,n-x+d)}\theta^{(x+c-1)}(1-\theta)^{(n-x+d-1)}, 0<\theta<1$$

In this framework the predictive preposterior distribution depending on data *x* is:

$$p(x,n) = \binom{n}{x} B(x+c,n-x+d)/B(c,d)$$

where *B* is a beta function. In this framework the ACC criterion becomes:

$$\sum_{x=0}^{n}\left\{\binom{n}{x}/B(c,d)\right\}\int_{a(x,n)}^{a(x,n)+l}\theta^{(x+c-1)}(1-\theta)^{(n-x+d-1)}\mathrm{d}\theta \geqslant 1-\alpha$$

Average Length Criterion

The ACC defines the sample size, fixing the coverage probability $(1-\alpha)$ of the HPD interval. The first step, in this case, is to find in data space the interval lengths $l'(x,n)$ satisfying this condition

$$\int_{a(x,n)}^{a(x,n)+l'(x,n)} f(\theta|x)\mathrm{d}\theta = 1-\alpha$$

The optimal sample size is the minimum integer which satisfies the condition:

$$\int_{x} l'(x,n)f(x)\mathrm{d}x \leqslant l$$

In this relation, l is the prespecified length. In this case, the left side of the equation is a mean of the lengths of the HPD intervals among data in the data space, weighted by the predictive distribution $f(x)$.

Considering the beta-binomial setting, the length *l'(x, n)* may be found by solving:

$$\int_{a(x,n)}^{a(x,n)+l'(x,n)} f(\theta|x,n,c,d)\mathrm{d}\theta = 1-\alpha$$

$$\int_{a(x,n)}^{a(x,n)+l'(x,n)} \frac{1}{B(x+c,n-x+d)}\theta^{(x+c-1)}(1-\theta)^{(n-x+d-1)}\mathrm{d}\theta = 1-\alpha$$

Once the candidate lengths in the data space have been found, the optimal sample size may be obtained by finding the minimum sample size such that:

$$\sum_{x=0}^{n} l'(x,n)p(x,n) \leqslant l$$

$$\sum_{x=0}^{n} l'(x,n)\binom{n}{x}B(x+c,n-x+d)/B(c,d) \leqslant l$$

The solution of this equation is not necessarily an HPD interval; for this reason, it is important to verify the conditions proposed in the literature to guarantee that a generic credibility interval is an HPD interval [17].

Worst Outcome Criterion

WOC criterion fixes both coverage probability $(1-\alpha)$ and length l in advance, ensuring a specified length and a minimum coverage among data x in data space.

The sample size may be found by choosing the minimum n, ensuring that:

$$\inf_{x\in X}\left\{\int_{a(x,n)}^{a(x,n)+l} f(\theta|x)\mathrm{d}\theta\right\} \geqslant 1-\alpha$$

Considering the case of the estimation for a single proportion, the relation on the left side of the equation becomes:

$$\int_{a(x,n)}^{a(x,n)+l} f(\theta|x,n,c,d)\mathrm{d}\theta$$

This integral cannot be minimized for all the values of *n, c, d* and *l,* and some conditions have to be identified to find a subset of data x^* as indicated in the literature [8]

$$\int_{a(x^*,n)}^{a(x^*,n)+l} f(\theta|x^*,n,c,d)\mathrm{d}\theta \geqslant 1-\alpha$$

The subset data x^* is defined by

$$x^* = \begin{cases} \dfrac{n+c+d+1}{2}-c \text{ or } \dfrac{n+c+d-1}{2}-c, & \text{if } n+c+d \text{ is odd and } n \geqslant |d-c| \\ \dfrac{n+c+d}{2}-c, & \text{if } n+c+d \text{ is even and } n \geqslant |d-c| \\ n & \text{if } 0 \leqslant n \leqslant |d-c| \end{cases}$$

Semiparametric approach for prior elicitation

The prior distribution for proportions may be derived in a semiparametric approach, eliciting an expert opinion as proposed by Bornkamp [13].

In this framework the prior distribution is obtained by performing optimization of a weighted sum of two components:

1. A term assessing the goodness of fit of a prior distribution among experts quantiles
2. The distance of the prior respect to a uniform uninformative distribution.

A uniform distribution has been considered as target prior, as suggested in the literature[13], in order to obtain a prior distribution which is a flexible compromise between a totally uninformative prior and an informative function adapted among expert quantiles.

The optimization problem proposed in this setting is based on the minimization of the Euclidean L^2 distance between functions $\int_{y_0}^{y_1}(1/(y_1-y_0)-f(y))^2 dy$, where $f(y)$ is a density function under the constraint $\int_{x_0}^{x_1} f(y)dy = 1$. This minimization problem is proportional to the Brier entropy defined for a range $[y_0, y_1]$.

The functional form of the probability density function among expert quantiles is approximated by a linear combination of B-splines with inner knots corresponding to specific boundaries. In this theoretical framework, F is a spline having m degree with a sequence of S inner knots $\lambda = (\lambda_{-m}, \dots, \lambda_{S+m+1})^T$. Some constraints have been imposed to guarantee that the linear combination is a density function.

1. The knots have to be ordered $\lambda_{-m} \leq \cdots \leq \lambda_{S+m+1}$
2. The spline has *(m + 1)* knots on each boundary point and S internal knots. The total number of knots on the boundaries is *2 (m + 1)*
3. The function F is defined on the domain $[y_0, y_1]$ and $F(y) = \sum_{i=-m}^{S} F_i N(y, m, \lambda_i)$
4. $F_i \in \mathbb{R}$ where $N(y, m, \lambda_i)$ is a normalized spline having a sequence of knots $\lambda_i = (\lambda_i, \dots, \lambda_{i+m+1})^T$

A control polygon connecting knots averages $\overline{\lambda_i}$, which determines the shape of the function.

Assuming to have *p*-elicited quantiles $y_{\alpha_1}, \dots, y_{\alpha_p}$ modeled by a linear combination of B-splines, the expert density function may be determined by optimizing this objective function:

$$\min_{F_{-m},\dots,F_S} \left\{ \sum_{i=1}^{p} \left(\alpha_i - F\left(y_{\alpha_i}\right) \right)^2 + \phi \int_{y_0}^{y_1} f(y)^2 dy \right\}$$

$$F_i \leq F_{i+1} \text{ for } i = -m, \dots, S-1$$

$$\textit{and } F_{-m} = 0, F_S = 1$$

where $F\left(y_{\alpha_i}\right)$ is the linear combination of B-splines (nonparametric component) stated among expert quantiles, and $f(y)$ is the density function of a uniform distribution (parametric component). Here, $\phi > 0$ is a balancing factor penalizing the distance between the function $F\left(y_{\alpha_i}\right)$ adapted to the expert quantiles and the uniform distribution in the domain $[y_0, y_1]$. Greater values of ϕ determine a prior distribution more similar to uniform, but instead, smaller values guarantee a posterior density which is better adapted among the expert quantile distribution.

This kind of optimization problem may be solved easily as addressed in the literature using a quadratic programming method [13]; moreover, additional constraints to the optimization problem may be imposed to obtain a monotone unimodal function [13].

As addressed in the literature [13] the degrees *m* of the spline may be defined between 2 and 5, and the inner knots may be defined on the expert's quantiles.

When no inner knots are defined (S = 0), the posterior distribution is a mixture of beta densities (a Bernstein polynomial basis) [18].

The balancing factor ϕ may be defined fixing an expected Δ error reflecting the distance between expert distribution and stated *p* quantiles.

$$\Delta = \sqrt{\frac{1}{p}\sum_{i=1}^{p}\left(\alpha_i - F\left(y_{\alpha_i}\right)\right)^2}$$

Having no information about the Δ error, default values are suggested in the literature [13].

Semiparametric B-splines approach for sample size estimation

The approaches developed in a parametric setting for binary data (ACC, ALC, and WOC) sample size estimation are extended incorporating a semiparametric B-splines prior obtained eliciting an expert's opinion.

B-Splines Average Coverage Criterion

A B-Splines Average Coverage Criterion (BSACC) sample size involves the same optimization problem as the ACC incorporating a different prior distribution.

$$\sum_{x=0}^{n} Pr\{\theta \in (a(x,n), a(x,n)+l)\}p(x,n) \geqslant 1-\alpha$$

In this framework, the coverage level may be reported as:

$$Pr\{\theta \in (a(x,n), a(x,n)+l)\} \propto \int_{a(\mathbf{x},n)}^{a(x,n)+l} \theta^x(1-\theta)^{(n-x)} f(\theta, m, S, \phi, y)\mathrm{d}\theta$$

In this relation $f(\theta, m, S, \phi)$ is the prior distribution obtained with the B-splines method, depending on *m* degrees of approximation, *S* inner knots and ϕ balancing factor and y_{α_i} expert quantiles

$$f(\theta, m, S, \phi, y) = \min_{F_{-m},\ldots,F_S}\left\{\sum_{i=1}^{p}\left(\alpha_i - F\left(y_{\alpha_i}\right)\right)^2 + \phi\int_{y_0}^{y_1} f(y)^2 dy\right\}$$

In this context, the predictive preposterior distribution depending on data *x* is:

$$p(x,n) = \int_{\Theta} \theta^x(1-\theta)^{(n-x)} f(\theta, m, S, \phi, y)\,\mathrm{d}\theta$$

The BSACC criterion becomes

$$\sum_{x=0}^{n} \int_{a(\mathbf{x},n)}^{a(x,n)+l} \theta^x (1-\theta)^{(n-x)} f(\theta, m, S, \phi, y) \mathrm{d}\theta \int_{\Theta} \theta^x (1-\theta)^{(n-x)} f(\theta, m, S, \phi, y) \mathrm{d}\theta \geqslant 1-\alpha$$

B-splines Average Length Criterion

The same minimization criterion has been provided for BSALC for a corresponding ALC.

Considering the B-splines prior distribution setting, the length *l'(x, n)* may be found by solving:

$$\int_{a(x,n)}^{a(x,n)+l'(x,n)} \theta^x (1-\theta)^{(n-x)} f(\theta, m, S, \phi, y) \mathrm{d}\theta = 1-\alpha$$

Once the candidate lengths in the data space have been found, the optimal sample size may be obtained by finding the minimum sample size such that:

$$\sum_{x=0}^{n} l'(x,n) \int_{\Theta} \theta^x (1-\theta)^{(n-x)} f(\theta, m, S, \phi, y) \mathrm{d}\theta \leqslant l$$

B-splines Worst Outcome Criterion

The B-splines Worst Outcome (BSWOC) criterion may be found by imposing the same constraints as indicated in the WOC method for parametric solutions, imposing that the prior be adapted to an expert opinion in a unimodal function [8]. Some conditions have to be identified to find a subset of data x^*[19]

$$\int_{a(x^*,n)}^{a(x^*,n)+l} f(\theta | x^*, m, S, \phi, y) \, \mathrm{d}\theta \geqslant 1-\alpha$$

Sample Size estimation procedure

A sample size estimation plan has been defined using a Generalized ACC, ALC and WOC estimation procedure (GACC, GALC, GWOC) useful to estimate the sample size considering different kinds of prior distributions defined in a parametric framework (ACC, ALC, WOC) or using semiparametric approaches based on B-splines defined on expert opinions (BSACC, BALC, BSWOC).

The method is generalized because it is based on HPD interval estimation among all data in the sample space, leading to consideration of a wide range of prior distributions, not only a conventional beta-binomial parametric solution.

The only constraint to consider is that the prior distribution considered in the computation has to be unimodal.

Prior Elicitation procedure

The design of a phase II clinical trial which aims to assess the effects of a drug on a binary safety endpoint in the pediatric population serves as the motivating example. At the time of writing, the authors are not entitled for confidentiality reasons to disclose all the data.

No objective data were available at the planning stage to define the prior. In this situation, an elicitation experiment has been conducted to obtain a prior distribution of the expert opinion.

An elicitation questionnaire has been submitted to obtain a prior distribution of the probability to observe an adverse event in children having a specific disease. All the experts involved in the research currently use the treatment in their clinical practice.

Eight opinions yi about the probability to observe an event in pediatric patients have been obtained by asking the experts the following question:

"*Based on your experience, what is the probability that in a patient aged 2 to 2, with a value of procalcitonin>1 μg / L, treated with a generic drug, has evidenced the presence of an event at 6 months from the acute episode?*"

The opinions provided by the experts, about the probability to observe an adverse event, are **y** = {0.30, 0.25, 0.15, 0.40, 0.30, 0.20, 0.20, and 0.30}. The mean of the provided opinions is 0.26, and the variance is 0.00625.

Informative, Low-Informative and Uninformative prior scenarios are considered for computation using parametric beta prior and a semiparametric solution.

Considering the beta parametric setting, the strength of prior influence on the final estimation has been defined using a power prior approach [20].

Different levels of penalization (discounting) may be provided on the expert opinion in order to perform sensitivity analysis on the prior choices.

The expert opinion may be included in the final computation using a $Beta(\alpha_i, \beta_i)$ prior where:

$$\alpha_i = \alpha_0 d_0 + 1$$

$$\beta_i = \beta_0 d_0 + 1$$

The α_0 and $\beta_0$values are the parameters obtained from the mean μ and variance σ^2 of the expert opinions using an inverse formula where:

$$\alpha_0 = \left[\left(\frac{1-\mu}{\sigma^2} - \frac{1}{\mu}\right)\mu^2\right] - 1$$

$$\beta_0 = \left[\alpha\left(\frac{1}{\mu} - 1\right)\right] - 1$$

The value d_0 defines the amount of expert information to be included in the final result. The discounting factor, otherwise, is defined as $(1 - \mathrm{d}_0) \times 100$ and represent the levels of penalization (discounting) on the expert opinion.

3. If $\mathrm{d}_0 = 0$, the data provided by the literature are not considered indicating a 100% discounting on prior information. According to this scenario, the prior is an uninformative $Beta(1,1)$ distribution.
4. If d_0=1 all the information provided by the experts is considered in the inference, indicating a 0% discounting of the expert opinion.

For the semiparametric sample size computation, the strength of influence of the expert opinion on the final result has been defined as varying the ϕ parameter.

In this general setting, 6 different scenarios were hypothesized for the prior computation (Figure 1)

Informative Priors

The expert opinions are used to obtain informative prior probability distribution in a parametric or semiparametric setting, considering:

1. A prior distribution $Beta(\alpha_i, \beta_i)$ with shape and scale obtained from the mean μ and variance σ^2 of the expert opinions, considering d_0=1 (0% discounting) corresponding to $Beta(8,22)$.
2. A B-splines semiparametric prior defined considering the inner knots located on the expert quartiles with *m=4* and a $\phi = 0.138$ derived from a $\Delta = 0.146$ as indicated in the literature [13].

Low-Informative Priors

Low-Informative priors have been defined in the computation considering:

1. $Beta(\alpha_i, \beta_i)$ with a d_0=0.5 (50% discounting) corresponding to $Beta(4.5,11.5)$.
2. A B-splines semiparametric prior with *m=4* degrees and ϕ=1.

Uninformative Priors

The uninformative priors have been compared with other scenarios respectively in parametric and semiparametric settings deriving following priors:

1. A prior distribution beta (1,1) with d_0=0 (100% discounting);
2. A B-splines semiparametric prior defined considering the inner knots located on the quartiles defined by an expert with *m=4* degrees and ϕ=45.

The optimal sample size has been found as the minimum integer ensuring a length of 0.2 with coverage equal to at least 0.95.

GACC GALC estimation

For each sample size among all likelihoods in data space, a Bayesian HPD interval has been estimated. All intervals with a length equal to 0.2 (or with coverage equal to 0.95 for ACC) have been selected.

Among the intervals, a weighted average of the coverage (or length for ALC) has been computed by weighting the posterior predictive distribution.

The optimal sample size is the minimum integer with coverage equal to at least 0.95 (or a length at most equal to 0.2 for ALC).

GWOC estimation

Considering WOC, a plan has been conducted considering different sample sizes performing a grid search around a frequentist sample size estimation.

For each sample size, among a subset of x^* likelihoods in data space, a Bayesian HPD interval has been estimated.

The values of expected success and failures c and d, for a semiparametric setting, have been found by drawing 1000 random samples by B-splines prior simulating 1000 binomial experiments. The median of the resampled success gives the prior expected successes c.

Frequentist sample size estimation

A frequentist estimate of sample size for a proportion has also been reported, for comparison purposes, considering a length of 0.2 and a confidence level of 0.95. The mean of the expert opinion (0.26) has been considered for the assumed point estimate.

Results

The frequentist estimated sample size is 75, in several cases higher compared to informative prior scenarios (Table 1).

Beta informative optimal sample sizes are similar across different estimation methods (GACC, GALC, GWOC), and smaller compared to designs obtained with other prior distributions as shown in Table 1. Generally, the sample sizes are more conservative for GWOC and smaller for GALC criterion (Table 1). Low-Informative sample sizes are a compromise between informative and

uninformative scenarios considering different estimation methods and semiparametric and parametric priors (Table 1).

The estimates provided for an uninformative prior estimation are, generally, higher compared to informative prior results (Table 1). Moreover, the sample size estimates provided for in semiparametric B-splines are generally comprised in the sample size derived with the parametric informative and uninformative scenario. A greater variability across results is evidenced for in the beta parametric scenarios compared to B-splines priors. Moreover, a 50% discounting factor ensures similar results compared to a Low-Informative B-splines prior scenario (ϕ=1) for ALC and WOC scenarios (Table 1).

By observing the pattern of the average coverage among all possible intervals in the sampling space, according to sample size (for a fixed length equal to 0.2), it is possible to evaluate a general increase in sample size as the average coverage level increases (Figure 2). The coverage is higher for all sample sizes for a beta informative prior distribution and is not much different for other scenarios; in this setting, semiparametric results are more similar to the beta informative scenario (Figure 2).

Considering the average length among all possible intervals ensuring coverage of 0.95, it is possible to show evidence that this value decreases with a decreasing sample size. Higher average length is observed for the uninformative beta scenario, while the lower average lengths are evidenced for the beta informative prior. The B-splines elicitation leads to an average length of HPD comprised between the values derived in both beta scenarios (Figure 3).

Considering the phase II study design, an ALC method using an informative B-splines prior has been selected among scenarios ensuring a sample size of 50 patients. The informative prior has been considered in order to take into account in the sample size computation information given by the experts about treatment effects, considering that experts are basically in agreement and have experience about treatment administration. The semiparametric prior has been adopted because it takes into account both the central tendencies of the prior distribution without completely discarding the tail of the expert opinion. Among sample methods, the more conventional ALC has been selected because 95% coverage is ensured in the data space scenario regardless of length [21].

The computational time needed to obtain the optimal sample size is higher compared with the original procedure proposed by Joseph [8] (less than one second for all methods ACC, ALC and WOC).

However, the computational burden associated with the proposed procedure is not overly high and, in every case, less than 159 seconds considering that, for B-splines solutions, the optimization problem may be solved with an easy and very fast convergence [13].

For GWOC scenarios, the computations have been performed considering only the data subspace x^* suggested by the literature[19]. For this reason, the computational time required to obtain the sample size is less than 5 seconds in every case.

Computations have been performed using the R 3.5.2 [22] System with HDInterval [23], SEL[13], SampleSizeBinomial[19] and LernBayes[24] packages.

Discussion

Limited sample size trials open the way to alternative statistical methods for data design and analysis in clinical trials. A Bayesian method may be useful in this framework when poor accrual problems are evident in the clinical research, leading to the inclusion of prior information about treatment effects, reducing the uncertainty of provided final estimates [25]. In some cases, there may be little objective evidence available, and the expert opinion may be used to elicit an informative prior [12].

The elicitation process may be conducted not only in parametric but also in semiparametric and nonparametric settings, leading to a more flexible approach to translating expert opinions into probability distributions [26].

At this point, it would be necessary to use a study design that could consider different approaches to the prior definition. For example, De Santis [27] highlights the importance of a flexible approach to prior distribution in a Bayesian study design considering the possibility of a sample size estimation method which takes into account different discounting factors defined on the priors, yet without comparing among parametric and nonparametric solutions.

The proposed GACC, GALC, and GWOC estimations procedures provide a pliable method to define study designs taking into account experts' opinions, possibly also using nonparametric approaches.

In comparison to a conventional ACC, ALC, and WOC, the proposed methods lead to inclusion in the sample size estimation of alternative prior distributions in comparison to a beta-binomial framework.

Considering the design of a phase II clinical trial, it is possible to observe that the experts may be more or less in agreement about the treatment effect size, or may have different experience or knowledge about therapy under evaluation. In this context, a more flexible design leads to tailor

prior distribution, and its capability to influence final results, according to the real expert's knowledge or agreement about the efficacy of the treatment.

According to the demonstrated motivating example, the experts are basically in agreement about the treatment effect and have experience about treatment administration in their clinical practice. In this framework, a more flexible approach is needed for prior definition which must be informative around the expert median but have also to take into account the tails of the expert opinion, leading to include more uncertainty in the study design compared to beta informative prior opinion and greater informativeness compared to a beta (1,1) prior.

Considering the results of computations, the estimations provided using a generalized approach show more extreme scenarios for sample sizes computed in the parametric framework. The smaller sample sizes are observed for the informative beta scenario; the semiparametric sample size is basically between sample sizes estimated in beta informative and uninformative settings. Less variability among the results is observed in varying the tuning parameter ϕ across B-splines prior choices compared to the beta scenario. Moreover, a $\phi = 1$ value ensures similar results with respect to a parametric solution with a 50% discount for ALC and WOC procedures.

Basically, the B-splines approach guarantees less extreme solutions compared to a beta prior, giving the possibility to achieve sample sizes comparable to results obtained with a beta prior having 50% discount, (especially for ACC and WOC).

All this means that the generalized method allows for planning a study design giving the possibility to consider different kinds of a priori distributions, more or less informative, or adapted to the opinions of experts. The choice among priors may depend on the degree of confidence in the available information.

Such confidence may depend on the degree of expert's belief in the treatment effect, or on the agreement among the experts. Moreover, the methods also give the possibility to use informative parametric methods, when objective information is available to derive the prior distributions.

The general framework of results gives solutions in term of sample size estimation similar to the parametric methods for sample sizes introduced by Joseph [19]. The solutions provided by GWOC are more conservative, leading to the simultaneous optimization of the HPD length and coverage.

The sample sizes provided by GACC and GALC are different, resulting in smaller sample sizes for GALC; the same pattern is observed comparing the ACC and ALC methods proposed by Joseph in a beta-binomial framework [19]. The choice among criteria appears somewhat arbitrary.

It is possible to consider ALC as more conventional because it fixes the coverage and HPD intervals are computed regardless of length [9]. In every case, the choice among criteria averaging over the predictive distribution of the data or considering the worst possible outcome depends on the degree of risk one is willing to take in the final inference [9].

More computational efforts are needed to derive the optimal sample size using the proposed method compared to the Joseph approach as the computations have been performed simulating the HPD among all likelihood in the overall data space (for ACC and ALC methods) and searching for the sample size around the frequentist estimate. However, the gain regarding flexibility in the study design is considerable, especially for a phase II study design and other studies conducted on small sample sizes.

The method may be easily extended considering also a continuous outcome or comparing binary or continuous outcomes across different groups, and a wide range of priors in parametric, semiparametric, and nonparametric frameworks.

The main limitation is that the choice among alternative priors requires evident computational efforts, especially if the posterior is not obtained in closed form or with an easy and fast convergence. More computational time may be required for study design providing comparisons among groups.

Conclusion

The generalized solutions to sample size definition gives the possibility to define a flexible Bayesian experimental design using different prior definitions. The approach is useful, especially in a clinical trial conducted on small sample sizes and when no objective evidence is available and may be indispensable to summarize an expert opinion.

The semiparametric sample size solutions are a compromise between the same estimates provided in a beta informative and uninformative setting ensuring not too different results, also varying the tuning parameter ϕ.

Performing a comparison across methods, GWOC is a more conservative solution. Instead, GALC gives smaller sample sizes compared to GWOC, and the same pattern is observed in the corresponding methods proposed in a beta-binomial setting.

References

[1] McNeish D. On using Bayesian methods to address small sample problems. Structural Equation Modeling: A Multidisciplinary Journal 2016;23:750–73.

[2] Hampson LV, Whitehead J, Eleftheriou D, Brogan P. Bayesian methods for the design and interpretation of clinical trials in very rare diseases. Statistics in Medicine 2014;33:4186–201. doi:10.1002/sim.6225.

[3] Administration UF and D. Guidance for the use of Bayesian statistics in medical device clinical trials. Maryland: US Food and Drug Administration 2010.

[4] Adcock C. A Bayesian approach to calculating sample sizes. The Statistician 1988:433–9.

[5] Pham-Gia T, Turkkan N. Sample size determination in Bayesian analysis. The Statistician 1992:389–397.

[6] Givon MM. Determination of optimal sample sizes in the beta-binomial brand choice model. Journal of Marketing Research 1980:58–62.

[7] Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Bayesian data analysis. vol. 2. CRC press Boca Raton, FL; 2014.

[8] Joseph L, Wolfson DB, Du Berger R. Sample size calculations for binomial proportions via highest posterior density intervals. The Statistician 1995:143–54.

[9] M'lan CE, Joseph L, Wolfson DB. Bayesian sample size determination for binomial proportions. Bayesian Analysis 2008;3:269–96.

[10] Quintana M, Viele K, Lewis RJ. Bayesian Analysis: Using Prior Information to Interpret the Results of Clinical Trials. Jama 2017;318:1605–6.

[11] Spiegelhalter DJ. Incorporating Bayesian ideas into health-care evaluation. Statistical Science 2004;19:156–74.

[12] Garthwaite PH, Kadane JB, O'Hagan A. Statistical methods for eliciting probability distributions. Journal of the American Statistical Association 2005;100:680–701.

[13] Bornkamp B, Ickstadt K. A note on B-splines for semiparametric elicitation. The American Statistician 2009;63:373–7.

[14] Oakley JE, O'Hagan A. Uncertainty in prior elicitations: a nonparametric approach. Biometrika 2007;94:427–41.

[15] Berchialla P, Zohar S, Baldi I. Bayesian sample size determination for phase IIA clinical trials using historical data and semi-parametric prior's elicitation. Pharmaceutical Statistics 2018.

[16] Lee JJ, Liu DD. A predictive probability design for phase II cancer clinical trials. Clinical Trials 2008;5:93–106.

[17] Box GE, Tiao GC. Bayesian Inference in Statistical Analysis. WISCONSIN UNIV MADISON DEPT OF STATISTICS; 1973.

[18] Goodman T, Sharma A. A modified Bernstein-Schoenberg operator. University of Dundee. Department of Mathematical Sciences; 1987.

[19] Joseph L, Wolfson DB, Du Berger R. Sample size calculations for binomial proportions via highest posterior density intervals. The Statistician 1995:143–54.

[20] Ibrahim JG, Chen M-H. Power prior distributions for regression models. Statistical Science 2000;15:46–60.

[21] JOSEPH L, Du Berger R, Bélisle P. Bayesian and mixed Bayesian/likelihood criteria for sample size determination. Statistics in Medicine 1997;16:769–81.

[22] R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2015.

[23] Meredith M, Kruschke J. HDInterval: highest (posterior) density intervals. R package version 0.1. 3. 2017.

[24] Albert J. LearnBayes: Functions for Learning Bayesian Inference (2008). R Package Version n.d.;2.

[25] Billingham L, Malottki K, Steven N. Small sample sizes in clinical trials: a statistician's perspective. Clinical Investigation 2012;2:655–7.

[26] Ghahramani Z. Bayesian non-parametrics and the probabilistic approach to modelling. Phil Trans R Soc A 2013;371:20110553.

[27] De Santis F. Using historical data for Bayesian sample size determination. Journal of the Royal Statistical Society: Series A (Statistics in Society) 2007;170:95–113.

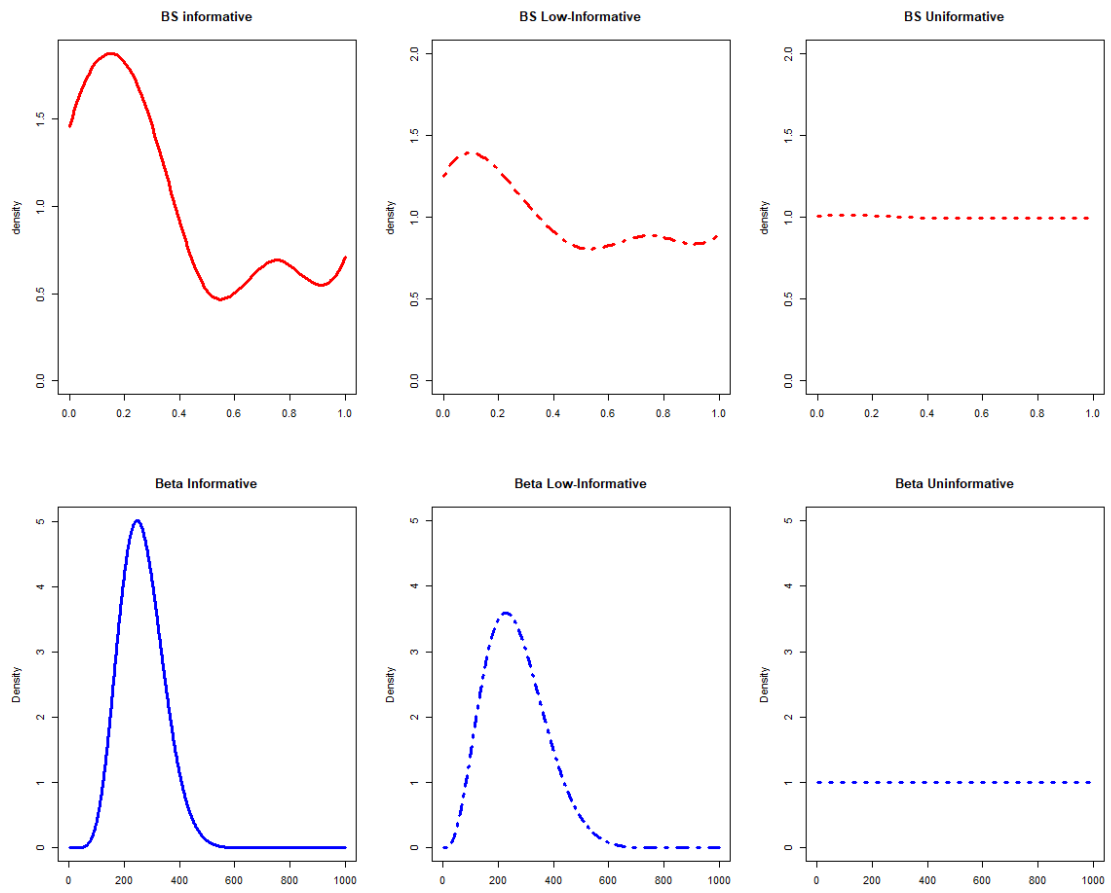

Figure 1 Prior distributions in a parametric and semiparametric setting considering Informative, Low-Informative and Uninformative scenarios. *BS stands for B-splines.*

Table 1 Optimal sample sizes defined following GACC, GALC, GWOC estimation methods using different prior distributions, length=0.2, coverage=0.95. BS stands for B-splines.

	GACC	GALC	GWOC
Beta Informative	43	42	45
Beta Low-Informative	59	53	76
Beta Uninformative	75	58	92
BS Informative	70	51	71
BS Low-Informative	76	54	77
BS Uniformative	77	56	86

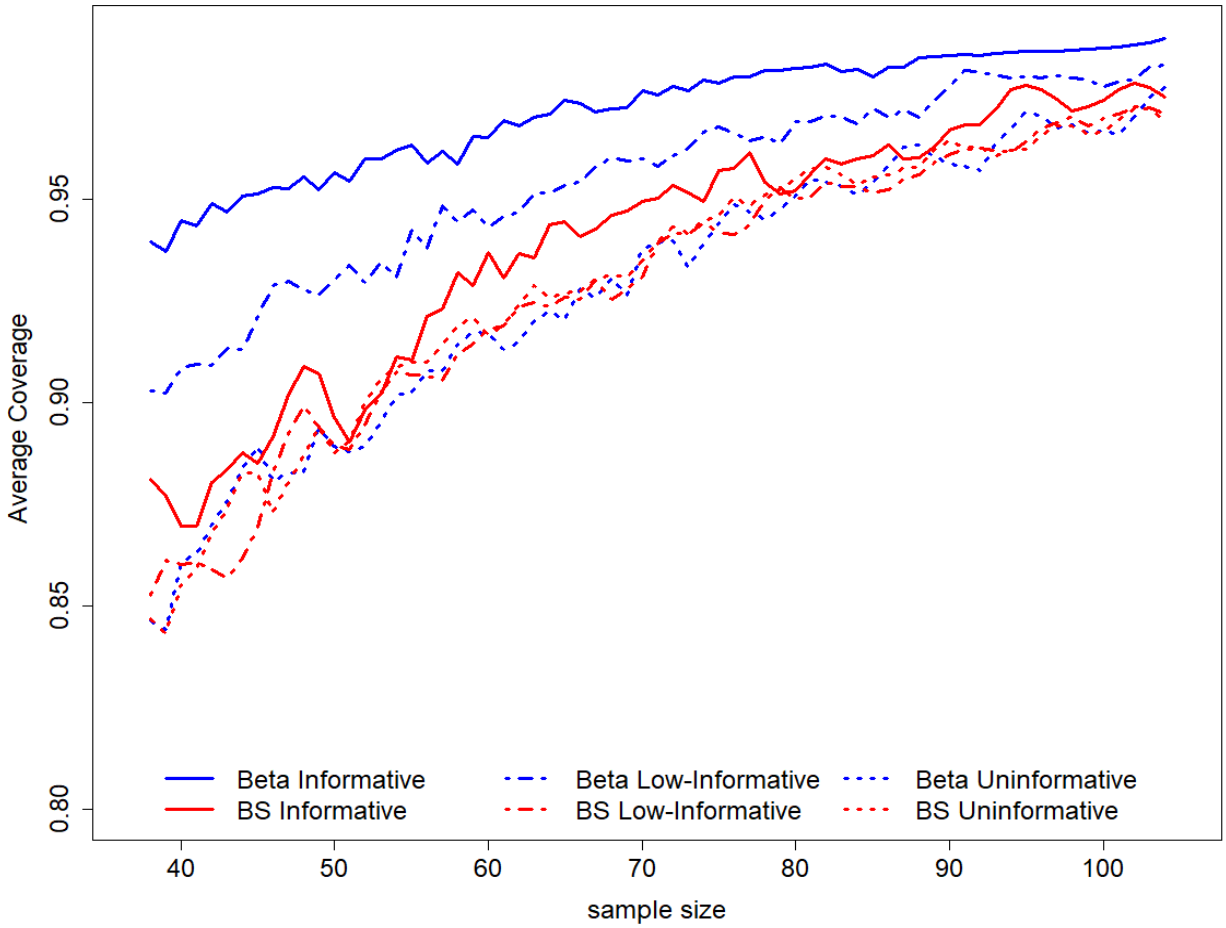

Figure 2 GACC estimation average coverage according to sample sizes for length=0.2. On the top side of the graph, the prior distributions used to perform the study design have been represented. BS stands for B-splines.

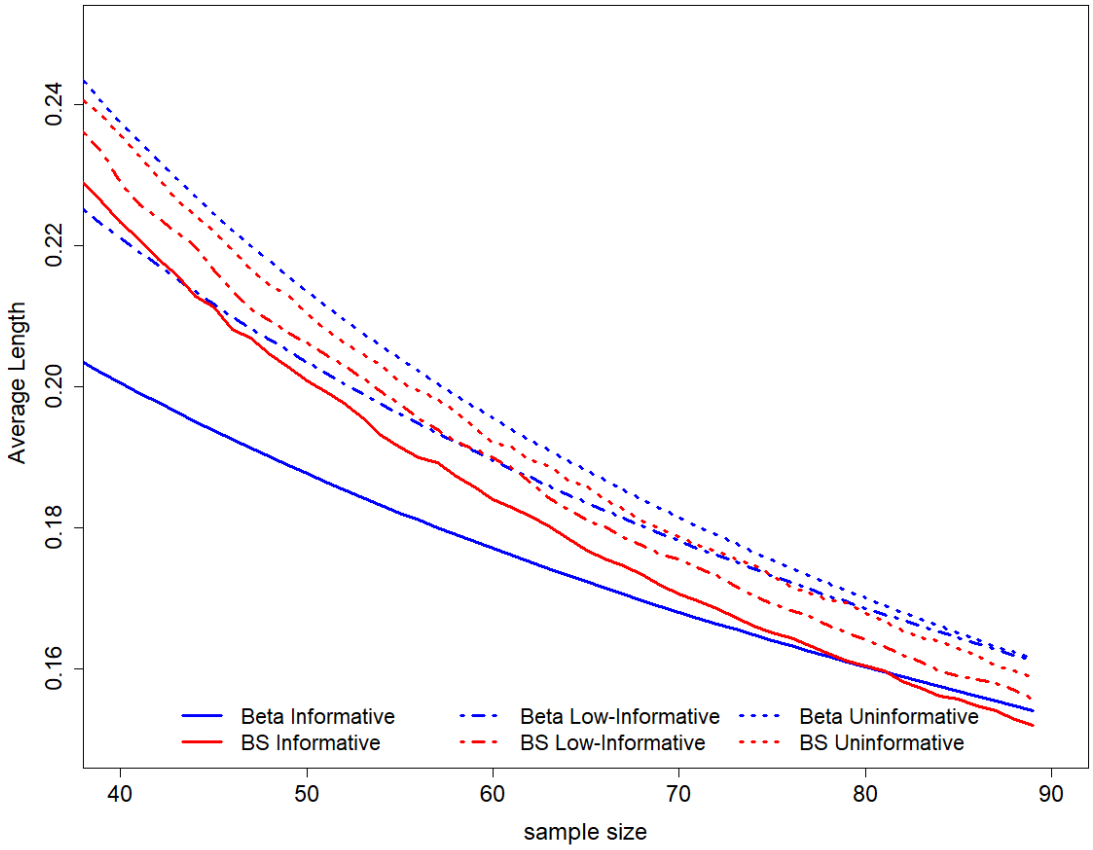

Figure 3 GALC estimation average length according to coverage equal to 0.95. On the top side of the graph, the prior distributions used to perform study design have been represented. BS stands for B-splines.

Conclusion

In the different clinical trial research contexts, like pediatric trials, several difficulties have been observed mainly in the enrollment process.
The Rescue trial is an example of a complex trial in pediatric research conducted in a poor accrual setting. The Bayesian analysis conducted on this trial allowed investigators to combine information provided by trial data with the evidence provided by the literature through the posterior distribution of the difference in proportion.
The first article (Chapter 1) evidenced the importance of the sensitivity analysis performed on prior choices. In fact, the posterior distribution of the absolute difference in proportions is highly influenced by the prior and is weakly influenced by the data, when using informative priors.
It follows that the correct definition of informative priors plays a fundamental role in a Bayesian trial especially if the study is conducted on limited sample sizes. Empirical evidence is not always available to derive objective prior; it follows that sometimes it is necessary to consider the expert opinion to obtain subjective prior distributions.

The second article, a literature review (Chapter 2), performed an analysis of the state-of-the-art of the prior elicitation methods in clinical trials. This research has shown that the last decade has seen increased the popularity of prior elicitation theme and the gap between theory and its application getting narrower and narrower. However, in clinical trial applications, conventional parametric approaches, to the prior elicitation, are the manly preferred solutions. It is widely assessed that the main limit of a parametric approach is to constrain expert belief into a pre-specified distribution, instead, not parametric and semi-parametric hybrid are more flexible methods.

A Bayesian trial has to be defined by the protocol, performing a sample size estimation taking into account expert opinion. The third article (Chapter 3) proposed a generalized version of the ACC, ALC, and WOC including, in the sample size estimation method, not only parametric approaches (Beta priors) but also semi-parametric priors based on B-Spline.
The application of the generalized method to a Phase II study design evidenced that the proposed approach is very pliable and suitable in this research context allowing to plan a study design considering the possibility to take into account of a wide range of a priori distributions, more or less informative, or adapted among the experts' opinion.

More computational efforts are needed to derive the optimal sample size compared to the conventional ACC, ALC and WOC approach.

However, the gain regarding flexibility in the study design is considerable, especially for phase II studies and other trials conducted on small sample sizes.

The method may be easily extended considering also continuous outcomes or comparing outcomes across different groups, replicating a similar simulation setting, considering a wide range of prior in parametric, semi-parametric, but also a not-parametric framework.

Appendix

Introduction to Bayesian Inference

Considering $X = \{x_1, \ldots, x_n\}$ data having sample size equal to n, the Bayes rule leads to incorporate in final inference two component in order to obtain a posterior distribution of treatment effect given data defined as $p(\theta \mid X)$:

$$p(\theta \mid X) = \frac{p(\theta)p(X \mid \theta)}{p(X)}$$

In this relation $p(X \mid \theta)$ is the Likelihood representing the information provided by the data, and $p(\theta)$ is the available knowledge about treatment effect translated into a prior distribution.

The prior distribution $p(\theta)$ may be defined using:

- Literature about treatment effect (Objective prior)
- Expert opinion about treatment effect (Subjective prior)

The quantity $p(X)$ is the predictive distribution of data defined as pre-posterior marginal distribution indicated as:

$$p(X) = \int_{\Theta} p\,(X, \theta) d\theta$$

Inference on the difference in proportion

A binomial experiment $X_i \sim \text{Bin}(n_i, \pi_i)$ may be defined on *i=1, 2* treatment groups, for which the sample sizes are n_i and the unknown probabilities to observe an event π_i may be expressed using a beta prior distribution.

$$\pi_i \sim Beta(\alpha_i, \beta_i)$$

The Beta distribution, having support defined between 0 and 1, is often used to model proportions and probabilities. The prior parameter α and β are derived by the expected number of event and failures.

In general, considering only one sample, assuming that the data are $X \sim \text{Bin}(n, \pi)$, placing a prior on π following a $\text{Beta}(\alpha, \beta)$ distribution, yields the following:

$$p(\pi|X) = \frac{p(X|\pi)p(\pi)}{\int p(X|\pi)p(\pi)d\pi}$$

$$\propto p(X|\pi)p(\pi)$$

$$\propto \prod_{i=1}^{n} \text{Bin}(n,\pi)\text{Beta}(\alpha,\beta)$$

$$\propto \prod_{i=1}^{n} \pi^{X_i}(1-\pi)^{n_i-X_i}\pi^{\alpha}(1-\pi)^{\beta}$$

$$\propto \pi^{\alpha+\sum_{i=1}^{n}X_i}(1-\pi)^{\beta+\sum_{i=1}^{n}n_i-\sum_{i=1}^{n}X_i}$$

In this case, $p(\pi|X)$ is proportional to another beta distribution and the posterior distribution for π is:

$$\pi|X \sim \text{Beta}(\alpha+\sum_{i=1}^{n}X_i\,,\beta+\sum_{i=1}^{n}n_i-\sum_{i=1}^{n}X_i)$$

Where $\sum_{i=1}^{n}X_i$ is the number of event and $\sum_{i=1}^{n}n_i$ is the sample size n.

In this framework, the posterior distribution for a difference in proportion $(\pi_1-\pi_2)$ may be computed using a resampling procedure where:

4. The event rate π_1^* may be resampled from $\pi_1|X_1$, which is the posterior distribution for group 1.
5. The event rate π_2^* may be resampled from $\pi_2|X_2$, which is the posterior distribution for group 2.

The posterior distribution for the parameters related to the difference in proportions $(\pi_1-\pi_2)$ may be obtained as the difference from the previously resampled distributions.