

Remote expert elicitation to determine the prior probability distribution for Bayesian sample size determination in international randomized controlled trials: Bronchiolitis in Infants Placebo Versus Epinephrine and Dexamethasone (BIPED) Study

Jingxian Lan

The Hospital for Sick Children

Amy C Plint

Children's Hospital of Eastern Ontario Research Institute

Stuart R Dalziel

The University of Auckland

Terry P Klassen

University of Manitoba

Martin Offringa

The Hospital for Sick Children

Anna Heath (✉ anna.heath@sickkids.ca)

Hospital for Sick Children <https://orcid.org/0000-0002-7263-4251>

Research Article

Keywords: Expert Elicitation, Bayesian Statistics, Randomised Controlled Trials, Sample Size Determination, Prior Probability Distribution, Trial Design

Posted Date: November 2nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-952484/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Title: Remote expert elicitation to determine the prior probability distribution for Bayesian sample size determination in international randomized controlled trials: Bronchiolitis in Infants Placebo Versus Epinephrine and Dexamethasone (BIPED) Study

Running Head: Expert elicitation and Bayesian SSD for BIPED

Word Count: 3992

Author List: Jingxian Lan, Amy C Plint, Stuart R Dalziel, Terry P Klassen, Martin Offringa, Anna Heath

Affiliations:

Jingxian Lan (phebe.lan@sickkids.ca): Child Health Evaluative Sciences, The Hospital for Sick Children, Toronto, Ontario, Canada

Amy C Plint (plint@cheo.on.ca): Division of Emergency Medicine, Children's Hospital of Eastern Ontario, Ottawa; Departments of Pediatrics and Emergency Medicine, University of Ottawa, Ottawa; Children's Hospital Research Institute, Ottawa, Ontario, Canada

Stuart R Dalziel (s.dalziel@auckland.ac.nz): Departments of Surgery and Paediatrics: Child and Youth Health, University of Auckland, and Children's Emergency Department, Starship Children's Hospital, Auckland, New Zealand.

Terry P Klassen (TKlassen@chrimg.ca): University of Manitoba, Winnipeg, Manitoba; Children's Hospital Research Institute of Manitoba, Winnipeg, Manitoba

Martin Offringa (martin.offringa@sickkids.ca): Child Health Evaluative Sciences, The Hospital for Sick Children, Toronto, Ontario, Canada; Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, Ontario, Canada; Division of Neonatology, The Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada

Anna Heath (anna.heath@sickkids.ca): Child Health Evaluative Sciences, The Hospital for Sick Children, Toronto, Ontario, Canada; Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada; Department of Statistical Science, University College London, London, United Kingdom.

Corresponding Author:

Anna Heath

The Hospital for Sick Children, Toronto

686 Bay Street, Toronto ON M5G 0A4

Phone: 416.813.7654 (x309500)

Email: anna.heath@sickkids.ca

Abstract

Background: Bayesian methods are increasing in popularity in clinical research. The design of Bayesian clinical trials requires a prior distribution, which can be elicited from experts. Current elicitation approaches either use face-to-face sessions or expert surveys. In diseases with international differences in management, the elicitation exercise should recruit internationally, requiring expensive face-to-face sessions or surveys, which suffer low response rates. To address this, we developed a remote, real-time elicitation exercise to construct prior distributions. These elicited distributions were then used to determine the sample size of the Bronchiolitis in Infants with Placebo Versus Epinephrine and Dexamethasone (BIPED) Study, an international randomized controlled trial in the Pediatric Emergency Research Network (PERN). The BIPED study aims to determine whether the combination of epinephrine and dexamethasone, compared to placebo, is effective in reducing hospital admission for infants presenting with bronchiolitis to the emergency department.

Methods: We developed a web-based tool to support the elicitation of the probability of hospitalization for infants with bronchiolitis. Experts participated in online workshops to specify their individual prior distributions, which were aggregated using the equal-weighted linear pooling method. The Average Length Criterion determined the BIPED sample size.

Results: Fifteen paediatric emergency medicine clinicians from Canada, USA, Australia and New Zealand participated in three workshops to provide their elicited prior distributions. The elicited probability of admission for infants with bronchiolitis was slightly lower for those receiving epinephrine and dexamethasone compared to supportive care in the aggregate distribution. There were substantial differences in the individual beliefs but limited differences between North America and Australasia. From this aggregate distribution, a sample size of 410 patients per arm results in an average 95% credible interval length of less than 9% and a relative predictive power of 90%.

Conclusion: Remote expert elicitation is a feasible, useful and practical tool to determine a prior distribution for international randomized controlled trials. Bayesian methods can then determine the trial sample size using these elicited prior distributions. The ease and low cost of remote expert elicitation means that this approach is suitable for future international randomized controlled trials.

Trial Registration: clinicaltrials.gov Identifier: NCT03567473

Key Words: Expert Elicitation, Bayesian Statistics, Randomised Controlled Trials, Sample Size Determination, Prior Probability Distribution, Trial Design

Background

Bayesian statistical methods use Bayes' theorem to combine information from observed data with previous evidence, characterised in a prior distribution, to make inferences about the parameters in a statistical model (1). Bayesian methods have become increasingly popular in clinical research as concern about frequentist methods had increased for several reasons (2). They can formally incorporate external evidence into the trial conclusions, rather than making definitive conclusions based on evidence from a single trial (3). They also provide a more natural interpretation of uncertainty (4), and easily allow for frequent monitoring and adaptive designs (5).

69 To take advantage of Bayesian methods, the sample size for the proposed trial must be determined.
70 Bayesian methods for sample size determination (SSD) have several advantages over frequentist SSD
71 methods. First, Bayesian SSD methods incorporate the statistical uncertainty that is inherent in the
72 estimates of key quantities (6). This contrasts to frequentist SSD methods where fixed values for several
73 key quantities, such as size and the target difference (7), must be specified and the required sample size
74 is highly sensitive to the values selected for these quantities (8). Secondly, frequentist SSD methods do
75 not consider clinicians' current beliefs about a treatment, meaning that trial results that contradict
76 strongly held beliefs are often not convincing enough to change clinical practice (9). Finally, sample sizes
77 calculated using frequentist methods are often hard to achieve or even infeasible in rare diseases (10).
78 In this setting, Bayesian SSD methods can reduce the required sample size by combining trial data with
79 other information, such as expert knowledge or earlier studies, to provide a similar level of scientific
80 certainty (11).

81 To utilize Bayesian SSD methods, a "prior distribution" must be defined to represent the currently
82 available evidence about the key parameters of interest (12). This prior distribution can be defined in
83 several ways, including using historical empirical data (13), expert knowledge or a combination of the
84 two (8). To use expert knowledge as the basis for a prior distribution, this knowledge must be converted
85 into a quantitative expression. This is commonly achieved through a structured "elicitation process" (14)
86 in which experts are assisted in converting their knowledge into a prior distribution through a series of
87 steps that are viewed as formal data acquisition processes based on validated methodologies (15).

88 Expert elicitation in clinical trials is becoming more frequent. A recent literature review identified 42
89 studies relating to clinical trial design and analysis from 460 studies discussing Bayesian prior elicitation
90 (16). Elicitation has been used, for example, in randomised controlled trials (RCTs) that compare
91 treatments for trauma resuscitation (17), bacterial corneal ulcers (18) and in rare diseases (19).

92 However, these elicitation studies required experts to meet in person, which can be difficult to arrange,

extremely expensive, especially in international studies, and is currently restricted due to the COVID-19 pandemic. An alternative approach to in-person meetings is to undertake a survey (20). However, survey-based elicitation exercises often have low response rates and only allow for limited assistance during the expert elicitation session (21). Furthermore, experts are not able to discuss and calibrate their beliefs, which is key to most elicitation frameworks (22,23).

As the goal of a RCT is to gather robust empirical evidence that could change clinical practice and health outcomes, the prior for the key parameters in an international RCT should robustly represent the beliefs of experts in all health systems where the trial results would be implemented. This representation is particularly important in diseases where there are regional (international) differences in clinical practice and presentation patterns. Therefore, an alternative, efficient, remote elicitation process is required to generate representative priors to support Bayesian SSD for international RCTs.

Bronchiolitis, a viral infection of the small and medium airways, and the most common reason for infants less than one year of age to be admitted to hospital in the developed world, is a disease with strong regional differences in clinical practice (24). Currently recommended management of bronchiolitis is predominantly the provision of parenteral fluids for hydration and oxygen for hypoxemia, called “supportive care” (25,26,27,28,29,30). Despite a lack of high-quality evidence, use of additional pharmacotherapy such as nebulized epinephrine, albuterol, hypertonic saline or oral corticosteroids varies by region, with an odds of use of any of these of 11.5 in Canada and 6.8 in the United States, compared to Australia and New Zealand (24). While the provision of pharmacotherapy is not supported by most guidelines, exploratory evidence suggests that the combination of inhaled epinephrine and oral corticosteroids has the potential to reduce hospital admission by a third in infants presenting to emergency departments (EDs) with bronchiolitis (31).

The Bronchiolitis in Infants with Placebo versus Epinephrine and Dexamethasone (BIPED) study is an international RCT comparing inhaled epinephrine and oral dexamethasone (a corticosteroid) to placebo in the treatment of infants presenting to EDs with bronchiolitis for the primary outcome of reducing admission into hospital, taking place in Canada, New Zealand, and Australia. Given the regional differences in bronchiolitis management and the geographical spread of BIPED sites, we set out to develop a remote, real-time elicitation exercise to overcome the limitations of the currently used elicitation exercises. From this exercise, we were able to provide a well-justified, representative prior to be used in the SSD and analysis of the BIPED study. This paper describes our novel approach to remotely elicit expert opinions for the BIPED study and the resulting Bayesian SSD.

Methods

The BIPED study

The BIPED study is a phase III, multi-centre, randomized, double-blind, placebo-controlled trial within the Pediatric Emergency Research Network (PERN) (32) to determine whether the combination of inhaled epinephrine and oral dexamethasone (EpiDex) is successful at reducing hospitalisation within the seven days following an initial presentation to an ED with bronchiolitis. The BIPED study is enrolling participants across 12 international sites; 6 in Canada, 3 in New Zealand and 3 in Australia. The study will enrol infants aged between 60 days and one year who present to the ED with an episode of wheezing or crackles, alongside signs of an upper respiratory tract infection during the peak season for Respiratory Syncytial Virus (RSV). The active treatment, to be compared with a placebo control, is two treatments of epinephrine (either via nebulisation (3 mg) or via metered dose inhaler and spacer (625 mcg)) given 30 minutes apart in the ED and two doses of once daily oral dexamethasone (0.6 mg/kg per dose, up to a maximum of 10mg). Participants will be randomised in 1:1 ratio to either the placebo or the EpiDex combination therapy. The BIPED study aims to provide the requested additional evidence

(33,34) after a previous study unexpectedly found that EpiDex had reduced symptoms sufficiently to decrease hospitalization within 7 days of an ED visit by one-third (31).

Research ethics approval

The BIPED study was approved by Health Canada and the local research ethics committee at each study site prior to enrollment. The remote elicitation exercise was approved by the Hospital of Sick Children research ethics committee. Implied consent was used for the remote elicitation exercise, meaning that by partaking in the elicitation exercise, the experts agreed that their data could be used for research purposes.

Designing the Remote Elicitation Exercise

Key Parameters and Clinical Setting. The primary outcome in the BIPED study is admission to hospital within 7 days following initial presentation to ED with bronchiolitis, which can be modelled using a binomial distribution. The key parameter of interest in the BIPED study is the probability of hospital admission within 7 days for each arm, placebo and EpiDex, denoted π_1 and π_2 , respectively. Beta distributions are commonly used to model beliefs about probabilities as the beta distribution is constrained between 0 and 1, has a flexible shape and is conjugate to the binomial likelihood (35). Thus, in our elicitation exercise, we assume that each expert's prior can be expressed as a beta distribution.

To enable the elicitation, we developed a clinical case study (see Supplementary Material) of an infant with bronchiolitis, who would meet the inclusion/exclusion criteria of the BIPED study, and was likely equivocal with respect to admission into hospital (i.e., EpiDex could potentially improve infant prognosis if prior beliefs supported benefit). Experts were then asked to determine the expected number of patients, out of 100, with characteristics similar to this patient who would be admitted to hospital within 7 days under two different treatment options (EpiDex and supportive care). The goal of the elicitation

exercise was to determine prior distributions for the BIPED Bayesian SSD and analysis. However, we decided that there was limited available expertise on the probability of admission under placebo and focussed on eliciting the probability of admission under supportive care. We then assumed that the outcomes under supportive care would be similar to placebo in our Bayesian SSD.

Developing an Online Elicitation Tool. Our remote elicitation exercise was based on the Sheffield Elicitation Framework (SHELF) methodology (17,23). Online tools have been developed to support the use of the SHELF framework (36) and we adapted these tools to support our elicitation about the number of hospitalizations for infants with bronchiolitis. For our remote elicitation exercise, we built a web-based interactive elicitation tool using R software and the shiny package (37,38) (<https://phebelan.shinyapps.io/Elicitation/>). In the online tool, experts were first asked to provide the lower and upper plausible values that subjectively described their beliefs about the number of infants with bronchiolitis who would be hospitalised within 7 days. We assumed that the lower and upper plausible values represented the limits of the 95% central credible interval in the beta distribution. Experts were provided their “Best” estimate for the number of hospitalizations, which we assumed was then the mode of the beta distribution. Within the interface, we restricted the value for the mode to be within the plausible interval. Using this method, we aim to prevent experts from anchoring to their initial selection and thereby underestimating uncertainty (22). Within the online tool, experts were provided with a real-time individual beta distribution plot and a quantitative summary of their beliefs to help adjust their estimates if the fitted beta distribution did not represent their beliefs (see Supplementary Material).

While the online tool supported the elicitation process, the Research Electronic Data Capture (REDCap) application collected the elicited distributions from each expert. REDCap is a web-based application designed to support secure data capture for research studies (39,40). Once we developed the online elicitation tool and REDCap database, we piloted the elicitation workshop three times internally (AP, SD,

TK, MO) to ensure clarity of expression, understanding and acceptability of the tool. We piloted these workshops remotely to ensure they could be delivered seamlessly and were an efficient use of experts' time.

Selecting the Experts. The BIPED study is being conducted in 12 sites across Canada, Australia and New Zealand. Therefore, we aimed to recruit experts from Canada, the United States, Australia and New Zealand to determine representative aggregate priors across the regions in the study, avoiding selection bias. To be eligible for the study, participants had to be (i) individuals identified as experts in bronchiolitis and its treatment and (ii) clinicians with experience in pediatric emergency medicine (PEM). Participants were excluded if they had extensive prior involvement with the BIPED study, i.e., serving as a site principal investigator. Potential participants were invited to volunteer to contribute by email. We aimed to recruit between 10 and 20 experts to ensure a breadth of experience in terms of geography and speciality (14,41).

Determining an Aggregate Prior Distribution. In elicitation, determining an aggregate prior distribution from the individual level distributions is viewed as a consensus formation process, in which the pooled prior distributions should fairly represent all individuals' beliefs (42). In our elicitation study, each expert $i = 1, \dots, N$ generates a prior distribution for each trial arm $j = 1, 2$,

$$p_{i,j} | \mathbf{x}_{i,j} \sim \text{Beta}(\alpha_{i,j}, \beta_{i,j}),$$

where $\mathbf{x}_{i,j} = (x_{i,j}^1, x_{i,j}^2, x_{i,j}^3)$ is the lower plausible value, mode and upper plausible value, respectively, from the expert elicitation process. These individual-level distributions are combined using the equal-weighted linear pooling method as it can reduce biases introduced by overoptimism and overconfidence (43). Thus, the pooled distributions will be equal to

$$\pi_j \sim \frac{1}{N} \sum_{i=1}^N p_{i,j} | \mathbf{x}_{i,j}$$

and will represent group's beliefs on the admission rate of infants with bronchiolitis under supportive care and EpiDex, respectively for $j = 1, 2$. We generated separate pooled distributions for each region and for each workshop to explore differences.

The Remote Elicitation Workshop

Pre-Workshop Materials. One week prior to the workshop, all participants were sent an email containing a study dossier to read before attending the workshop. The goal of this dossier was to introduce the concept of an elicitation exercise and the currently available literature on treatments for bronchiolitis (22). Our study dossier included a published elicitation study (17) and four published studies presenting the use of epinephrine and/or dexamethasone as a treatment for bronchiolitis (31,44,45,46). The goal of including a previous elicitation study was to introduce the experts to the concept of elicitation, while the other studies were included to complement the experts' knowledge with the current literature.

Remote Expert Elicitation Workshop. We conducted three remote elicitation workshops using Zoom, a cloud-based video conferencing platform (47). Three facilitators from the BIPED study team with statistical and medical expertise attended each workshop. The workshop began with an introduction to the BIPED study and the rationale of Bayesian statistics. To familiarise experts with the elicitation procedure, an example using our online elicitation tool was then shown. Experts were then asked to provide their personal beliefs about the chance of hospitalisation for the patient identified in the case study.

The elicitation exercise was structured over two rounds with a group discussion between the two rounds (22,23). In the first round, experts used our online elicitation tool to provide their individual prior distribution for the probability of hospitalisation with supportive care and EpiDex. The facilitator (JL) then generated a deidentified boxplot (shown in Figure S1) to display all the individual-level priors and support the group discussion. The group discussion allowed the experts to adjust and calibrate their

responses but did not aim to reach a consensus (22). Thus, the group discussion began with the facilitator interpreting the individual boxplots before the experts were encouraged to share their beliefs and discuss their thoughts around the observed variations in beliefs across experts. When the group discussion no longer resulted in an exchange of information, the facilitator would manage the discussion and to help promote critical thinking (48). Following the group discussion, experts were again asked to use the online elicitation tool to characterise their beliefs and these results then generated the individual prior distribution to be pooled.

Following the Remote Elicitation Workshop. Following the completion of all three workshops, the experts were sent the pooled distributions for the probability of hospital admission with supportive care and EpiDex. The experts were also sent the workshop-specific pooled distribution for each workshop and their own individual distributions for comparison and were invited to provide comments, if they had any.

Bayesian Sample Size Determination

To determine the sample size in the BIPED study, we use the average length criterion (ALC) for Bayesian SSD (49). This method selects the smallest sample size for which the average length of a specified posterior credible interval is below a given threshold. The ALC uses a preposterior analysis where the length of the posterior credible interval is estimated across the range of potential studies, as estimated by the prior-predictive distribution of the data (49). To achieve this, we simulated the probability of hospitalisation within 7 days under the two treatments based on the priors from the expert elicitation exercises using a binomial likelihood. These simulated data were combined with our aggregated prior to determine the posterior for the two probabilities of hospitalisation, using Markov chain Monte Carlo (MCMC) methods. We then calculated the 95% high density posterior credible interval for the difference in the probability of admission across the two treatments, placebo and EpiDex. We estimated the

average posterior credible interval length for sample sizes between 400 and 630 using 1500 simulations from the prior-predictive distribution and 5000 simulations from the posterior. We selected the sample size for which the ALC is below 0.09.

In the BIPED study, we will declare that EpiDex is superior to placebo if the posterior probability that the probability of hospitalisation under EpiDex is greater than the probability of hospitalisation under placebo exceeds a threshold λ ;

$$P(\pi_1 < \pi_2) > \lambda$$

To select the threshold λ , we simulated data assuming the probability of hospitalisation is equal to 0.35 for both EpiDex and placebo and selected a threshold λ such that $P(\pi_1 < \pi_2) > \lambda$ in at most 5% of the simulated studies. For the fixed value of λ , we then computed the frequentist power of study by computing the proportion of simulated studies with $P(\pi_1 < \pi_2) > \lambda$ when the probabilities of hospitalisation are 0.35 and 0.27 for placebo and EpiDex, respectively, representing a target difference of 8%. Finally, we compute the relative predictive power of the decision rule, defined as the proportion of simulated studies from the prior predictive distribution for which $P(\pi_1 < \pi_2) > \lambda$, standardised by the prior probability that EpiDex is superior to placebo. These three calculations were based on 8000 simulated trials with 5000 simulations from the posterior. All Bayesian analysis were performed using JAGS through R (38,50).

Results

Elicitation Workshop

Baseline Characteristics

We invited 25 PEM clinicians from Canada, the United States, Australia and New Zealand to participate in our three remote elicitation workshops. In total, 15 of these experts agreed to participate in the

study; 9 from North America (NA) and 6 from Australia and New Zealand (ANZ). The three workshops contained 5 (2 NA; 1 ANZ), 4 (4 NA) and 6 (3 NA; 3 ANZ) participants, respectively. Table 1 displays the baseline characteristics for these 15 experts. Experts from NA had more experience treating bronchiolitis with epinephrine and dexamethasone, separately and combined. However, most experts do not currently use either treatment in their routine practice.

Treatment	All	North America	Australasia
Number of responses	15	9	6
Has past Experience treating patients with: n (%)			
Epinephrine	10 (67)	9(100)	1 (17)
Dexamethasone	2 (14)	1(12)	1 (17)
Epinephrine and dexamethasone	5 (34)	4 (45)	1 (17)
Currently treating patients with: n (%)			
Epinephrine	4 (27)	4 (45)	0 (0)
Dexamethasone	0 (0)	0 (0)	0 (0)
Epinephrine and dexamethasone	0 (0)	0 (0)	0 (0)

Table 1: Baseline characteristics of experts, by region of practice.

Prior distributions

Figure 1 displays the individual prior distributions for the two probabilities of hospitalisation, with supportive care on the left and EpiDex on the right. The individual responses from both rounds were highly varied, both in terms of the central tendency of the distributions across individuals and the width of the plausible interval within individuals. Although most experts believe that the probability of

admission for infants with bronchiolitis is slightly lower for those receiving EpiDex compared to supportive care. Figure 2 displays the pooled prior distributions from all experts for each round in the elicitation workshop. In both rounds, the pooled prior distributions show a slight reduction in the probability of admission for infants with bronchiolitis who are treated with EpiDex. However, experts were less certain about the size of this reduction in the second round, demonstrating that the group discussion led the experts to be more conservative.

We explore the pooled prior distributions separately for each workshop (Figure S2) and across the two regions (Figure S3) in the supplementary material. The prior distributions demonstrate that similar beliefs about the probability of hospitalisation are held in NA and ANZ. However, the aggregate distributions were different for each workshop, likely due to the diversity of our experts and the limited number of individuals in each workshop.

Figure 1: Individual level elicited prior distributions for hospitalisation probability under a) supportive care (left), or b) treatment with the combination of epinephrine and dexamethasone (EpiDex, right). Each line depicts the distribution scored by an individual participant (n=15). Distributions for first elicitation round on top; second round at bottom.

Figure 2: Pooled elicited prior distributions for hospitalisation probability under a) supportive care (Supportive, solid black line), or b) treatment with the combination of epinephrine and dexamethasone (EpiDex, dashed red line). Distributions for first elicitation round top; second round bottom.

Bayesian Sample Size Determination

We computed the average length of the 95% high design posterior credible interval for the difference in admission probability between the two arms (Figure 3). From these results, we specify a sample size of 410 participants per arm for the BIPED study ensuring the average 95% credible interval is shorter than 9%. Adjusting for an expected 5% loss to follow up, the total sample size of the BIPED study is 432 per arm. The average 95% credible interval would be less than 8% if the BIPED study recruits 610 participants per arm.

With 410 participants per arm, we select $\lambda = 0.99$ as the threshold for declaring that EpiDex is superior to placebo. With this threshold, we incorrectly conclude that EpiDex is superior to placebo when there is no effect in 4.6% of the simulated trials (Type 1 error). This threshold then results in correctly concluding that EpiDex is superior in 81% of the simulated trials with a target difference of 8% and a relative predictive power of 90%.

Figure 3: Average 95% posterior credible interval length for “admission probability difference” between placebo and EpiDex plotted across the BIPED clinical trial sample sizes increasing between 400 and 630 in increments of 5 (solid black line). Average Length Criterion (ALC) thresholds of 0.09 and 0.08 are plotted as dashed black lines (see text).

Discussion

We developed a remote elicitation framework, which offers a practical and convenient method for expert elicitation. Expert belief, elicited using this framework, can then form the basis of a Bayesian SSD and analysis for an international RCT, where the prior distribution should represent the diverse beliefs across the regions enrolling patients. Our remote framework allowed us to practically obtain diverse opinions by running a synchronous online exercise with a reasonably large number of diverse experts.

330 We enrolled 15 experts from 4 countries (Canada, United States, New Zealand and Australia) within a
331 relatively short time frame, on a limited budget, under COVID-19 related travel restrictions and
332 determined a pooled prior distribution that represents the diversity of perspectives in an international
333 trial. As our elicitation exercise involved a relatively short time commitment, we were able to achieve
334 high response rates, resolving issues seen with survey-based elicitation (21). Finally, we were able to
335 hold multiple elicitation workshops assisted by a facilitator to further broaden the range of experts who
336 could attend.

337 Another advantage of our remote elicitation framework, compared to survey-based elicitation methods,
338 is that we were able to have real-time facilitation and a group discussion. This facilitated expert
339 interaction and allowed us to identify issues within the workshops. As can be seen by the differences
340 between the distributions between the two rounds, the group discussion was critical in calibrating the
341 experts' beliefs. In particular, the experts raised external factors that would influence the decision to
342 admit an infant with bronchiolitis, such as hospital resources and family circumstances. Experts also
343 shared their thoughts and clarifications related to the design of the elicitation exercise and their
344 understanding of the elicitation task. We were also able to respond to any technical issues and ensure
345 that all enrolled experts were able to provide responses.

346 The biggest challenge we encountered was scheduling the workshops to maximise attendance.
347 Challenges included large differences in time zones between the countries and accommodating the shift
348 patterns of practicing PEM clinicians working in the ED. We decided to run multiple workshops so a
349 greater number of experts could participate and aimed to include experts from each region in each
350 workshop and ensure there were enough participants to allow a fruitful group discussion. While we
351 were largely successful, we found the scheduling of these workshops to be a significant challenge and
352 highly recommend inviting a higher number of experts than required as some schedules may be
353 incompatible, especially when working across multiple time zones.

A strength and limitation of our remote elicitation exercise is time taken for the workshop. Each workshop was scheduled for 90 minutes and the experts were invited to read five manuscripts before attending the workshop as preparation – estimated to take another 90 minutes. This minimal time commitment, compared to day-long meetings and travel, allowed us to recruit a range of experts to our study and was key to enrolling practicing PEM physicians. However, the set 90 minutes meeting-slot did limit the time available for presenting the theory behind elicitation, which could have impacted the quality of our elicited prior distribution.

Conclusions

To overcome challenges associated with standard methods for trial SSD and analysis and to enable successful application of Bayesian methods, we developed a remote elicitation framework that offers a comprehensive, practical, affordable approach to obtaining prior distributions for a Bayesian analysis of an international RCT, where the current state of knowledge about the key parameters across the jurisdictions where the trial results will be implemented should be incorporated into the analysis. This prior distribution can be used to determine the appropriate sample size for a proposed Bayesian analysis of the completed RCT. Thus, our proposed remote elicitation process promotes the use of Bayesian methods in randomized controlled trials.

List of Abbreviations

ANZ: Australia and New Zealand; BIPED: Bronchiolitis in Infants Placebo Versus Dexamethasone Study; REDCap: Research Electronic Data CAPture; EpiDex: Combination therapy of nebulized epinephrine and dexamethasone; ED: Emergency department; NA: North America; PEM: Pediatric Emergency Medicine

Declarations

377 **Ethics approval and consent to participate:** We have obtained primary ethics approval from
378 the Hospital for Sick Children Research Ethics Board with reference number: 1000075048. The elicitation
379 study used implied consent in line with our ethics approval. No patients were involved.

380 **Consent for publication:** Not applicable

381 **Availability of data and materials:** Our consent procedures do not allow for data sharing due to the
382 small number of participants in the study and the ease of identifying these individuals. The pooled
383 distributional forms are available from the corresponding author on reasonable request.

384 **Competing Interests:** The authors report no competing interests. The authors alone are responsible for
385 the writing and content of this article.

386 **Funding:** This work is funded through an Innovative Clinical Trials Multi-year Grant from the Canadian
387 Institutes of Health Research (funding reference number MYG-151207; 2017 - 2020), as part of the
388 Strategy for Patient-Oriented Research and in partnership with the Alberta Children's Hospital Research
389 Institute (Calgary, Alberta), Centre Hospitalier Universitaire Sainte-Justine (Montreal, Quebec),
390 Children's Hospital Research Institute of Manitoba (Winnipeg, Manitoba), CHEO Research
391 Institute (Ottawa, Ontario), Hospital for Sick Children Research Institute (Toronto, Ontario), Stollery
392 Children's Hospital (Edmonton, Alberta), Research Manitoba (Winnipeg, Manitoba), University of
393 Western Ontario (London, Ontario), and the Women and Children's Health Research Institute
394 (Edmonton, Alberta). SRDs time was supported by Cure Kids New Zealand. ACP is supported in part by a
395 Tier I University of Ottawa Research Chair. TPK is supported by the CRC in Clinical Trials. AH is supported
396 by the CRC in Statistical Trial Design.

397 **Author Contributions:**

398 JL and AH designed the elicitation workshop. SRD, TPK, MO and ACP piloted the elicitation workshop and
399 drafted the clinical case. JL, SRD, TPK, ACP and AH facilitated the elicitation workshop. JL analysed the

data. JL and AH drafted the article. SRD, TPK, MO and ACP offered substantive revisions. All authors conceived the elicitation study and read, edited, and approved the final manuscript.

Acknowledgements: We thank all members of the KidsCAN PERC Innovative Pediatric Clinical Trials BIPED Study Group and the KidsCAN PERC Innovative Pediatric Clinical Trials Methods Core for assistance in developing the protocol and statistical analysis plan for the BIPED trial. We acknowledge the KidsCAN PERC Innovative Pediatric Clinical Trials team and our parent advisors (Serena Hickes, Kurt Schreiner, Julie Leung) who provided valuable input on the study design and documents. We thank the Pediatric Emergency Research Canada (PERC) network of health care professionals and the KidsCAN Trials Network for their contribution and support to this project and pediatric clinical research in Canada. Finally, we thank all participants in the remote elicitation workshop, Trevor Kuang, Antonia Stang, Robert Porter, James Chamberlain, Vikram Sabhaney, Jo Cole and those who wished to remain anonymous.

References

1. van de Schoot R, Depaoli S, King R, Kramer B, Märtens K, Tadesse MG, et al. Bayesian statistics and modelling.. *Nat Rev Methods Primers*. 2021; 1(1): <https://doi.org/10.1038/s43586-020-00001-2>.
2. Berry DA. Bayesian clinical trials. *Nature reviews Drug discovery*. 2006; 5(1): 27-36.
3. Spiegelhalter DJ, Freedman LS, Parmar MK. Bayesian approaches to randomized trials. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. 1994; 157(3): 357-387.
4. Jack Lee J, Chu CT. Bayesian clinical trials in action. *Statistics in medicine*. 2012; 31(25): 2955-2972.
5. Berry SM, Carlin BP, Lee JJ, Muller P. *Bayesian adaptive methods for clinical trials* Boca Raton, FL: CRC Press; 2010.
6. Adcock CJ. Sample size determination: a review. *Journal of the Royal Statistical Society: Series D (The Statistician)*. 1997; 46(2): 261-283.
7. Cook JA, Julious SA, Sones W, Hampson LV, Hewitt C, et al.. DELTA2 guidance on choosing the target difference and undertaking and reporting the sample size calculation for a randomised controlled trial. *BMJ*. 2018; 363: k3750.

8. Joseph L, Du Berger R, Bélisle P. Bayesian and mixed Bayesian/likelihood criteria for sample size determination. *Statistics in medicine*. 1997; 16(7): 769-781.
9. Chaloner K, Rhamé FS. Quantifying and documenting prior beliefs in clinical trials. *Statistics in medicine*. 2001; 20(4): 581-600.
10. 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(3): p.e0120981.
11. Choy SL, O'Leary R, Mengersen K. Elicitation by design in ecology: using expert opinion to inform priors for Bayesian statistical model. *Ecology*. 2009; 90(1): 265-277.
12. Cao J, Lee JJ, Alber S. Comparison of Bayesian sample size criteria: ACC, ALC, and WOC. *Journal of statistical planning and inference*. 2009; 139(12): 4111-4122.
13. De Santis F. Using historical data for Bayesian sample size determination. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. 2007; 170(1): 95-113.
14. Dias LC, Morton A, Quigley J. *Elicitation: The Science and Art of Structuring Judgement* Cham, Switzerland: Springer; 2018.
15. Estévez RA, Mardones FO, Álamos F, Arriagada G, Carey J, Correa C, et al. Eliciting expert judgements to estimate risk and protective factors for Piscirickettsiosis in Chilean salmon farming. *Aquaculture*. 2019; 507: 402-410.
16. Azzolina D, Berchialla P, Gregori D, Baldi I. Prior Elicitation for Use in Clinical Trial Design and Analysis: A Literature Review. *International journal of environmental research and public health*. 2021; 18(4): 1833.
17. Jansen JO, Wang H, Holcomb JB, Harvin JA, Richman J, Avritscher E, et al. Elicitation of prior probability distributions for a proposed Bayesian randomized clinical trial of whole blood for trauma resuscitation. *Transfusion*. 2020; 60(3): 498-506.
18. 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(6): 407-413.
19. Ramanan AV, Hampson LV, Lythgoe H, Jones AP, Hardwick B, Hind H, et al. Defining consensus opinion to develop randomised controlled trials in rare diseases using Bayesian design: An example of a proposed trial of adalimumab versus pamidronate for children with CNO/CRMO. *PLoS One*. 2019; 14(6): p.e0215739.
20. 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(1): 63-70.

21. Grigore B, Peters J, Hyde C, Stein K. EXPLICIT: a feasibility study of remote expert elicitation in health technology assessment. *BMC medical informatics and decision making*. 2017; 17(1): 1-10.
22. Gosling JP. SHELF: the Sheffield elicitation framework. In *Elicitation*. Cham, Switzerland: Springer; 2021. p. 61-93.
23. O'Hagan A, Oakley J. SHELF: The Sheffield Elicitation Framework. Sheffield, UK;; Accessed: July 7, 2021.
24. Schuh S, Babl FE, Dalziel SR, Freedman SB, Macias CG, Stephens D, et al. Practice variation in acute bronchiolitis: a pediatric emergency research networks study. *Pediatrics*. 2017; 140(6): e20170842.
25. Babl F, Borland M, Ngo P, Acworth J, Krieser D, Pandit S, et al. Paediatric Research in Emergency Departments International Collaborative (PREDICT): first steps towards the development of an Australian and New Zealand research network. *Emergency Medicine Australasia*. 2006; 18(2): 143-147.
26. Ricci V, Nunes VD, Murphy MS, Cunningham S. Bronchiolitis in children: summary of NICE guidance. *Bmj*. 2015; 350: h2305.
27. Ralston SL, Lieberthal AS, Meissner HC, Alverson BK, Baley JE, Gadomski AM, et al. Clinical practice guideline: the diagnosis, management, and prevention of bronchiolitis. *Pediatrics*. 2014; 134(5): e1474-e1502.
28. Friedman JN, Rieder MJ, Walton JM, Canadian Paediatric Society, Acute Care Committee. Bronchiolitis: recommendations for diagnosis, monitoring and management of children one to 24 months of age.. *Paediatrics & child health*. 2014; 19(9): 485-491.
29. Baumer JH. SIGN guideline on bronchiolitis in infants. *Archives of Disease in Childhood-Education and Practice*. 2007; 92(5): ep149-ep151.
30. Quinonez RA, Schroeder AR. Safely doing less and the new AAP bronchiolitis guideline. *Pediatrics*. 2015; 135(5): 793-795.
31. Plint AC, Johnson DW, Patel H, Wiebe N, Correll R, Brant R, et al. Epinephrine and dexamethasone in children with bronchiolitis. *New England Journal of Medicine*. 2009; 360(20): 2079-2089.
32. Klassen T, Dalziel S, Babl F, Benito J, Bressan S, Chamberlain J, et al. The Pediatric Emergency Research Network (PERN): A decade of global research cooperation in paediatric emergency care. *Emergency Medicine Australasia*. 2021;(doi: 10.1111/1742-6723.13801).
33. Hartling L, Fernandes RM, Bialy L, Milne A, Johnson D, Plint A, et al. Steroids and bronchodilators for acute bronchiolitis in the first two years of life: systematic review and meta-analysis. *Bmj*. 2011; 342: d1714.
34. Wainwright C. Acute viral bronchiolitis in children-a very common condition with few therapeutic options. *Paediatric respiratory reviews*. 2010; 11(1): 39-45.

35. Bolstad WM, Curran JM. Introduction to Bayesian statistics Hoboken, NJ: John Wiley & Sons; 2016.
36. Morris DE, Oakley JE, Crowe JA. A web-based tool for eliciting probability distributions from experts. *Environmental Modelling & Software*. 2014; 52: 1-4.
37. Chang W, Cheng J, Allaire J, C. S, Schloerke B, Xie Y, et al. shiny: Web Application Framework for R. R package version 1.6.0. Vienna, Austria;; 2021.
38. R Core Team. R: A language and environment for statistical computing. Vienna, Austria;; 2020.
39. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, et al. The REDCap consortium: Building an international community of software platform partners. *Journal of biomedical informatics*. 2019; 95: 103208.
40. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of biomedical informatics*. 2009; 42(2): 377-381.
41. European Food Safety Authority. Guidance on expert knowledge elicitation in food and feed safety risk assessment. *EFSA Journal*. 2014; 12(6): 3734.
42. Dietrich F, List C. Probabilistic Opinion Pooling. In *The Oxford Handbook of Probability and Philosophy*. Oxford, UK: Oxford University Press; 2016. p. 1-27.
43. Johnson SR, Tomlinson GA, Hawker GA, Granton JT, Feldman BM. Methods to elicit beliefs for Bayesian priors: a systematic review. *Journal of clinical epidemiology*. 2010; 63(4): 355-369.
44. Corneli HM, Zorc JJ, Mahajan P, Shaw KN, Holubkov R, Reeves SD, et al. A multicenter, randomized, controlled trial of dexamethasone for bronchiolitis. *New England Journal of Medicine*. 2007; 357(4): 331-339.
45. Florin TA, Plint AC, Zorc JJ. Viral bronchiolitis. *The Lancet*. 2017; 389(10065): 211-224.
46. Kirolos A, Manti S, Blacow R, Tse G, Wilson T, Lister M, et al. A systematic review of clinical practice guidelines for the diagnosis and management of bronchiolitis. *The Journal of infectious diseases*. 2020; 222(7): S672-S679.
47. Zoom Video Communications. Zoom Cloud Meetings. Accessed: August 9, 2021;; 2021.
48. Hemming V, Burgman MA, Hanea AM, McBride MF, Wintle BC. A practical guide to structured expert elicitation using the IDEA protocol. *Methods in Ecology and Evolution*. 2018; 9(1): 169-180.
49. Joseph L, M'LAN CE, Wolfson DB. Bayesian sample size determination for binomial proportions. *Bayesian Analysis*. 2008; 3(2): 269-296.

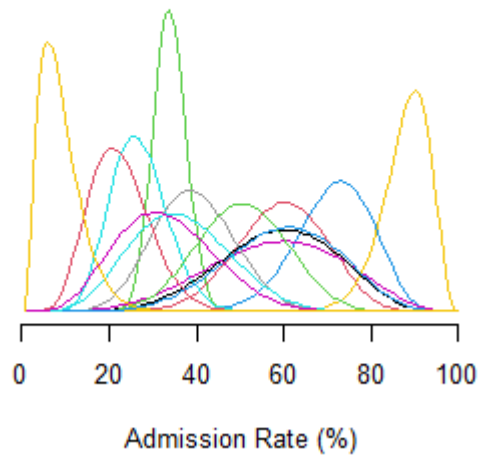
50. Plummer M. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. Proceedings of the 3rd international workshop on distributed statistical computing. 2003; 124(125.10): 1-10.

414

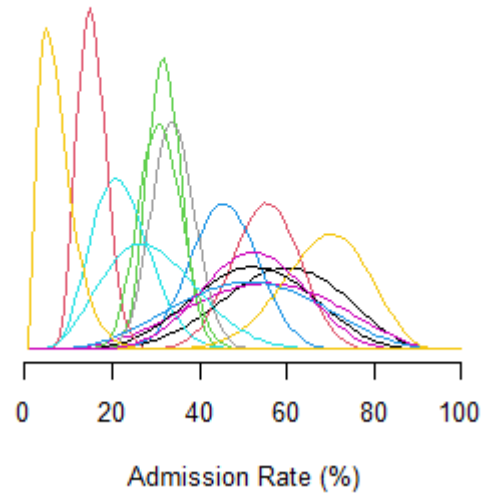
415

Figures

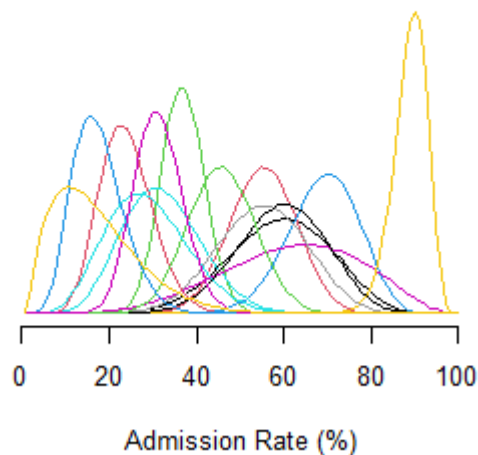
Individual prior distribution for the probability of hospitalization (Supportive Care) - Round 1



Individual prior distribution for the probability of hospitalization (EpiDex) - Round 1



Individual prior distribution for the probability of hospitalization (Supportive Care) - Round 2



Individual prior distribution for the probability of hospitalization (EpiDex) - Round 2

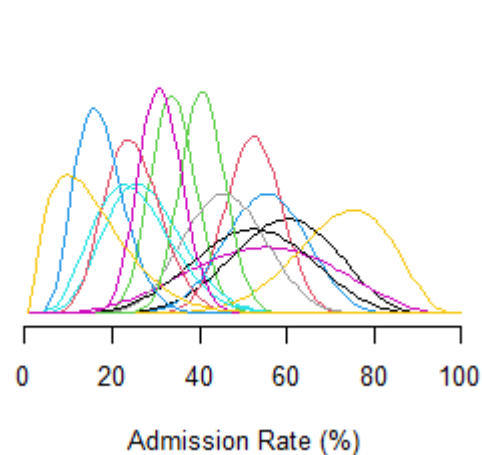


Figure 1

Individual level elicited prior distributions for hospitalisation probability under a) supportive care (left), or b) treatment with the combination of epinephrine and dexamethasone (EpiDex, right). Each line depicts the distribution scored by an individual participant (n=15). Distributions for first elicitation round on top; second round at bottom.

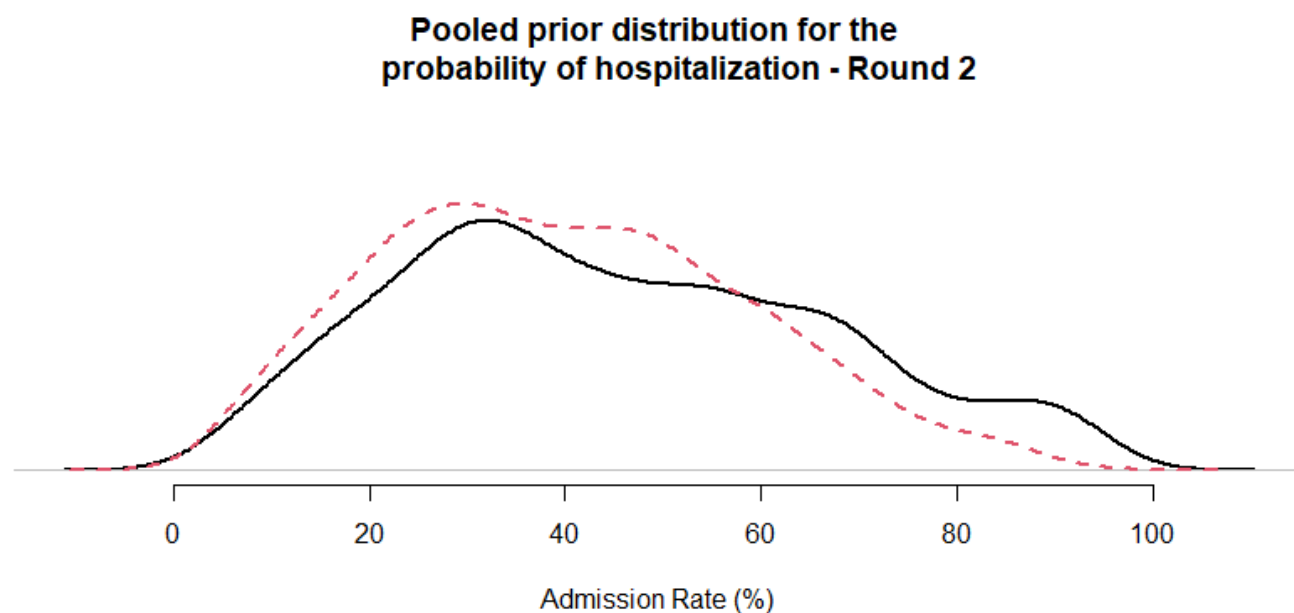
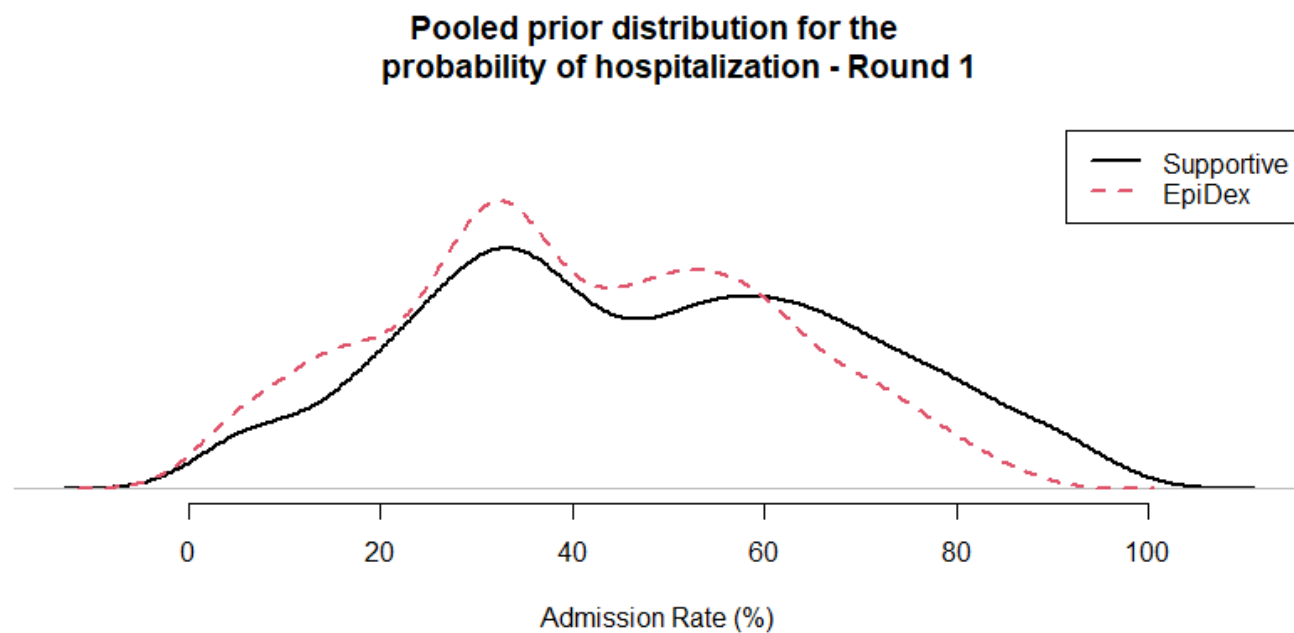


Figure 2

Pooled elicited prior distributions for hospitalisation probability under a) supportive care (Supportive, solid black line), or b) treatment with the combination of epinephrine and dexamethasone (EpiDex, dashed red line). Distributions for first elicitation round top; second round bottom.

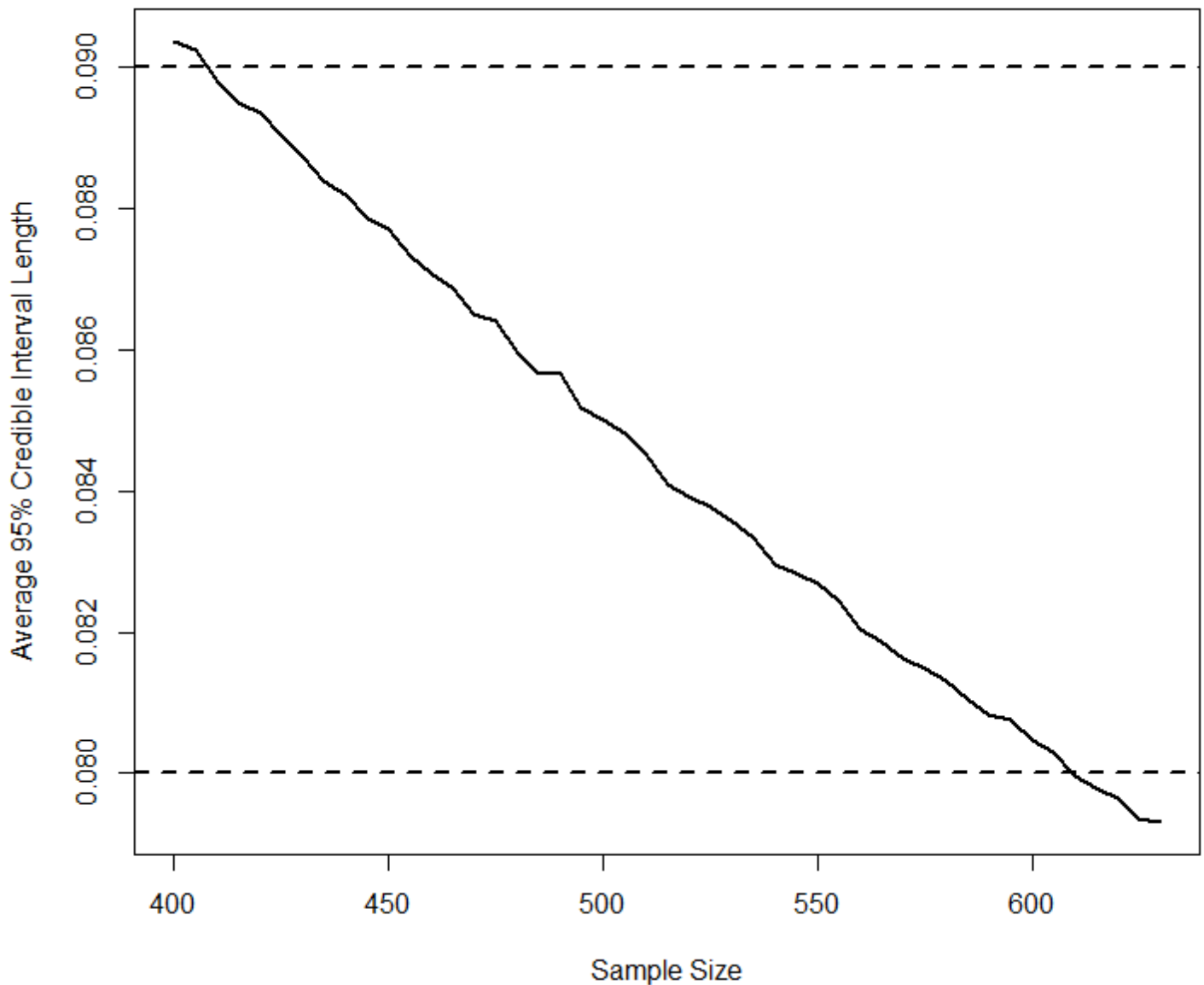


Figure 3

Average 95% posterior credible interval length for “admission probability difference” between placebo and EpiDex plotted across the BIPED clinical trial sample sizes increasing between 400 and 630 in increments of 5 (solid black line). Average Length Criterion (ALC) thresholds of 0.09 and 0.08 are plotted as dashed black lines (see text).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryMaterialV4.docx](#)