Statistics in Biopharmaceutical Research

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/usbr20

# A New Centralized Statistical Monitoring Method for Detecting Atypical Distribution of Qualitative Variables in Multicenter Randomized Controlled Trials

Serge Niangoran, Antoine Barbieri, Anani Badje, Valérie Journot, Gérard Kouame, Olivier Marcy, Xavier Anglaret & Amadou Alioum

**To cite this article:** Serge Niangoran, Antoine Barbieri, Anani Badje, Valérie Journot, Gérard Kouame, Olivier Marcy, Xavier Anglaret & Amadou Alioum (12 Nov 2024): A New Centralized Statistical Monitoring Method for Detecting Atypical Distribution of Qualitative Variables in Multicenter Randomized Controlled Trials, Statistics in Biopharmaceutical Research, DOI: 10.1080/19466315.2024.2404631

**To link to this article:** https://doi.org/10.1080/19466315.2024.2404631

STATISTICS IN BIOPHARMACEUTICAL RESEARCH
2024, VOL. 00, NO. 0, 1–9
https://doi.org/10.1080/19466315.2024.2404631



# A New Centralized Statistical Monitoring Method for Detecting Atypical Distribution of Qualitative Variables in Multicenter Randomized Controlled Trials

Serge Niangoran[a,b,c], Antoine Barbieri[a], Anani Badje[a,b,c], Valérie Journot[a,b], Gérard Kouame[c], Olivier Marcy[a,b], Xavier Anglaret*[a,b], and Amadou Alioum*[a]

[a]Univ. Bordeaux, National Institute for Health and Medical Research (INSERM) UMR 1219, Bordeaux Population Health, Bordeaux, France; [b]Research Institute for Sustainable Development (IRD) EMR 271, Bordeaux, France; [c]Programme PACCI, Abidjan, Côte d'Ivoire

**ABSTRACT**

Centralized statistical monitoring (CSM) detects clinical trial centers in which the distribution of a variable is atypical compared to its distribution in other centers. Most proposed CSM methods concern quantitative variables. Here we propose a new hierarchical Bayesian beta-binomial (HBBB) method for categorical variables and report the results of a simulation study assessing the performance of the method and of an application study using a real database to assess its usefulness. In the simulation study, sensitivity exceeded 90% when the sample size in the atypical center ($N_a$) was $\geq$20 and the difference in the proportion of events between the atypical center and the other centers ($\delta$) was $\geq$0.4; when $N_a$ was $\geq$40 and $\delta$ $\geq$0.3; and when $N_a$ was $\geq$150 and $\delta$ $\geq$0.2. Specificity exceeded 90% when $N_a$ was $\geq$150 in all scenarios, and remained between 75% and 90% when $N_a$ was lower. In the application study, the method detected two centers in which $N_a$ was 50 and 200, and $\delta$ was 0.12 and 0.04, respectively. The performance of the HBBB method was similar to that proposed by competing approach. The modeling is easy and specificity is good in many scenarios with a limited sample size.

**ARTICLE HISTORY**
Received September 2023
Accepted August 2024

**KEYWORDS**
Bayesian approach; Beta-binomial modeling; Categorical variable; Centralized statistical monitoring; Multicenter clinical trial

## 1. Introduction

Data monitoring is essential to both ensure patient safety and high-quality scientific evidence. This is an essential but potentially burdensome task, leading International drug regulatory bodies, including the Food and Drug Administration (FDA) and the European Medicines Agency (EMA), to recommend a risk-based approach to monitoring multicenter clinical trial data (Agency 2013; Center for Drug Evaluation and Research 2022). This approach involves centralized monitoring of data using new information and communication technologies, which include centralized statistical monitoring (CSM). CSM uses statistical tests or procedures to identify whether the distribution of a variable in one center differs from its distribution in other centers (Buyse et al. 1999). When an atypical pattern is detected, it can trigger further investigation, including on-site monitoring.

Most CSM methods proposed in the literature concern quantitative variables (Pogue et al. 2013; Desmet et al. 2014; de Viron et al. 2022). Few methods have been specifically developed to deal with categorical variables. Desmet et al. (2017) proposed a method based on beta-binomial distribution for the detection of an atypical center in terms of the distribution of a categorical variable, and reported that the method performed well (Desmet et al. 2017). Their modeling was inspired by the work of Chuang-Stein (1993) and Kleinman (1975) on the use of a beta-binomial model to summarize medical information in different observed groups.

In this article, a new hierarchical Bayesian beta-binomial (HBBB) method is proposed to detect an atypical center in terms of the distribution of a categorical variable. The approach is a combination of the Desmet approach, using the beta-binomial distribution, and the approach of Hatayama and Yasui (2020), who used a hierarchical Bayesian method for finite Gaussian mixture models.

The Bayesian approach we propose is hierarchical in the sense that we use hyperpiors on the parameters of the beta distribution, assigning them a distribution that takes into account the heterogeneity and complexity of the data (Kruschke and Vanpaemel 2015).

Examples of binary categorical variables on which this modeling can relate concern the incidence of cardiac toxicity of any grade in 15 clinical studies on an anticancer agent (Chuang-Stein 1993), or the proportion by center of missing data on the CD4 count collected in the clinical trial (Desmet et al. 2017).

We first describe the HBBB method. We then present the results of a simulation study to evaluate the performance of the method in terms of sensibility and specificity. We finally present the results of an application study using a real database to illustrate the practical use of the method.

**CONTACT** Serge Niangoran bessekon.niangoran@u-bordeaux.fr Inserm U1219, Université de Bordeaux, 146 rue Léo Saignat, 33076 Bordeaux Cedex, France.
*These authors contributed equally.

Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/SBR.

## 2. Material and Method

### 2.1. Description of the HBBB Method

#### 2.1.1. Beta-Binomial Distribution

We consider a multicenter clinical trial with M centers of size $N_i (i = 1, 2, \ldots, M)$ in which an event is recorded using a binary variable. $Y_i$ is the discrete random variable corresponding to the number of events in center $i$, with a binomial distribution, that is $Y_i \sim \text{Bin}(N_i, p_i)$, where $p_i$ is the probability of occurrence of the event. In the context of multicenter clinical studies, a binomial distribution cannot adequately describe the additional variation when $p_i$ varies, thus, we consider a beta-binomial model as proposed by Desmet et al. (2017).

Frequently used in Bayesian inference, the beta-binomial distribution is a discrete probability distribution with finite support, corresponding to a process of Bernoulli draws where the probability of occurrence of an event is random (following a beta distribution) (Lee and Sabavala 1987; Everson and Bradlow 2002). More precisely, $Y_i \sim \text{Bin}(N_i, p_i)$, where $p_i$ is a random variable with a beta distribution:.

$$\pi\left(p_i|\alpha,\beta\right) = \text{Beta}\left(\alpha,\beta\right) = \frac{p_i^{(\alpha-1)}(1-p_i)^{(\beta-1)}}{B(\alpha,\beta)} \quad (1)$$

where $B(\alpha,\beta) = \int_0^1 t^{(\alpha-1)}(1-t)^{(\beta-1)}dt$ and $\alpha > 0, \beta > 0$ are two shape parameters that govern how the underlying probabilities of events $p_i$ are distributed. If we assume that the variable $Y_i$ is in a beta-binomial distribution $BB(N_i, \alpha, \beta)$, this approximates the binomial distribution when parameters $\alpha and \beta$ are arbitrarily large. Another parameterization $BB(N_i, \mu, \rho)$ is possible for the beta-binomial distribution (as in the simulation study section below). While parameters $\alpha, \beta$ are the natural parameters for beta distribution, the usual parameters for beta-binomial distribution are the location (mean) parameter

$$\mu = \frac{\alpha}{\alpha+\beta}(0 < \mu < 1)$$

considered as a proportion, and the overdispersion parameter

$$\rho = \frac{1}{\alpha+\beta+1}(0 < \rho < 1).$$

When $\rho$ tends to 0, the beta-binomial distribution approximates the binomial distribution.

#### 2.1.2. The Hierarchical Bayesian Beta-Binomial (HBBB) Model

In this article a hierarchical Bayesian approach is proposed for estimating the beta-binomial $BB\left(N_i, \alpha, \beta\right)$ model parameters $\alpha, \beta$ from the whole data $(N_i, y_i)$ observed in the M centers, where $y_i$ is the observed number of events recorded in center $i$, and $N_i$ is the number of participants in center $i$. The proposed Bayesian approach consists of three hierarchical levels:

$$\begin{cases} \text{Level1: } y_i|p_i \sim \text{Bin}(N_i, p_i)\text{Samplingmodel} \\ \text{Level2: } p_i|\mu,\rho \sim \text{Beta}\left(\alpha = \left(\frac{1}{\rho}-1\right)\mu, \beta = \left(\frac{1}{\rho}-1\right)(1-\mu)\right)\text{Definitionofpriors} \\ \text{Level3: } \mu \sim \text{Beta}(a,b)\text{and}\rho \sim \text{Beta}(c,d)\text{Definitionofhyperpriors} \end{cases} \quad (2)$$

where $a, b, c$, and $d$ are fixed values for the beta distribution bounds.

The advantage of this parameterization using $\mu$ and $\rho$ instead of $\alpha$ and $\beta$, is that in the Bayesian approach, we can introduce a prior information on $\mu$ that is easily interpretable in terms of proportion of events, but also on $\rho$ in terms of overdispersion. Note that for non-informative priors, we will consider.

$$\mu \sim \text{Beta}\left(1,1\right)$$
$$\rho \sim \text{Beta}\left(1,1\right)$$

the choice of beta distribution parameters for $\mu$ and $\rho$ for informative priors should reflect the external or expert information we have on the proportion of events in the centers and the level of overdispersion.

Compared to a frequentist estimation of the beta-binomial model, one of the advantages of our Bayesian approach is the potential incorporation of prior information into our model through defining the prior distributions of the model parameters. This information about parameter values can come from expert knowledge (e.g., the expected proportion in centers should be 10%) or historical data. To facilitate understanding and explicit incorporation of this information, we redefined the prior distributions of the beta distribution parameters in terms of location (mean) and overdispersion parameters rather than the natural parameters of the distribution. For inference on model parameters, and for the sake of simplicity, it is more practical to return to the natural parameters $\alpha$ and $\beta$ of the beta distribution for the expression of the likelihood.

The sampling model likelihood of the data can be expressed as

$$L\left(y|p\right) = \prod_{i=1}^{M} f(y_i|p_i) = \prod_{i=1}^{M} \binom{N_i}{y_i} p_i^{y_i}\left(1-p_i\right)^{N_i-y_i} \quad (3)$$

where the structure of the model is hierarchical and is defined in (2).

In the Bayesian approach, the statistical inference or prediction is based on the posterior distribution of parameters given data. According to Bayes' rule, the posterior distribution is defined as

$$\pi(p|y) = \frac{f(y|p)\pi(p)}{f(y)}$$

where $f\left(y\right) = \int f(y|p)\pi(p)dp$ is the marginal distribution of the data.

Since $f(y|p) = \binom{N}{y} p^y(1-p)^{N-y}$ and $\pi(p) = \frac{p^{\alpha-1}(1-p)^{\beta-1}}{\text{B}(\alpha,\beta)}$, the marginal distribution can be expressed as

$$f\left(y\right) = \frac{\binom{N}{y}\text{B}(\alpha+y, \beta+N-y)}{\text{B}(\alpha,\beta)},$$

and the posterior probability distribution as

$$\pi\left(p|y\right) = \frac{p^{\alpha+y-1}\left(1-p\right)^{\beta+N-y-1}}{\mathrm{B}\left(\alpha+y,\beta+N-y\right)}.$$

As a result, the posterior distribution is also Beta $\left(\alpha+y,\beta+N-y\right)$. That is,

$$p_i|y_i \sim \text{Beta}\left(\alpha+y_i,\beta+N_i-y_i\right) \quad (4)$$

This result is a classic example of conjugate prior, since the prior in the beta-binomial model is a Beta, and the posterior is also a Beta. To estimate $\alpha$ and $\beta$, methods for sampling Monte Carlo simulations by Markov chains (MCMC) were used. MCMC sampling methods use a combination of the Gibbs sampler (for updates) and the Metropolis-Hastings (M-H) algorithm to construct samples of parameters $\alpha$ and $\beta$ from their posterior distribution. From the posterior sample of the $\alpha$ and $\beta$, the median values were retained, leading to a distribution of the posterior probabilities of the beta distribution.

### 2.2. Procedure for Detecting Atypical Center Using the HBBB Approach

Using the HBBB model, the distribution of the number of expected events was calculated from the posterior probability distribution for each clinical center. A Bayesian inference was then used to decide whether a center $i$ was atypical or not, by estimating the credibility interval at the $100\times\left(1-q\right)\%$ threshold associated with the number of predicted events $\left[\tilde{y}_i^{[q/2]};\tilde{y}_i^{[1-q/2]}\right]$, that is the region of higher posterior density events in the center (Robert 2007; Hespanhol et al. 2019). The credibility interval was such that:

$$P\left(y_i\in\left[\tilde{y}_i^{[q/2]};\tilde{y}_i^{[1-q/2]}\right]\right) = \int_{\tilde{y}_i^{[q/2]}}^{\tilde{y}_i^{[1-q/2]}}\binom{N_i}{y_i} \times\frac{\Gamma(\alpha+y_i)\Gamma(\beta+N_i-y_i)}{\Gamma(\alpha+\beta+N_i)}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}dy_i = 1-q. \quad (5)$$

Where $\tilde{y}_i^{[q/2]}$ and $\tilde{y}_i^{[1-q/2]}$ are respectively the lower and upper limits of the credibility interval for the difference between observed and predicted values by the posterior distribution applied to each center $i$. Thus, we used the quantiles of order q/2 and (1 − q/2) in order to obtain the (1 − q)% of higher posterior density (HPD) value of the observation in the investigative center.

The procedure for detecting an atypical center using the HBBB approach can be summarized as follows:

- Step 1: Estimation of HBBB model defined in (2) using Bayesian approach.
- Step 2: Get estimated parameter values $(\hat{\alpha},\hat{\beta})$ of $(\alpha,\beta)$ from their posterior sample.
- Step 3: Assuming a common distribution p for all centers, generate the L-sample p=$(p^{(1)},p^{(2)},\ldots,p^{(L)})\ldots)$ such as

$$p^{(l)} \sim \text{Beta}(\hat{\alpha},\hat{\beta}).$$

- Step 4: For each center i (i=1,…,M), generate the L-sample $\widehat{Y}_i=(\tilde{y}_i^{(1)},\tilde{y}_i^{(2)},\ldots,\tilde{y}_i^{(L)})$, of $Y_i$ such as $Y_i^{(l)} \sim \text{Bin}\left(N_i,p(l)\right)$ for $l=1,\ldots,L$.
- Step 5: Conclude that center $i$ is an atypical center if the observation $y_i,\notin\left[\tilde{y}_i^{[q/2]};\tilde{y}_i^{[1-q/2]}\right]$, otherwise the center $i$ is considered as non-atypical,

where $\tilde{y}_i^{[q/2]}$ is the quantile of order q ? ]0,1[ of sample $\widehat{Y}_i=(\tilde{y}_i^{(1)},\ldots,\tilde{y}_i^{(L)})$ and the number L is defined by the numbers of Markov chains and the iterations.

## 3. Simulation Study to Test HBBB Method Performance

### 3.1. Overview

A multicenter clinical trial was simulated, with M centers of the same size $N_i(i=1,2,\ldots,M)$, to observe the occurrence of an event characterized by a binary variable for each participant in the trial. The outcome of interest Y was the number of events observed in each center.

The simulated distribution of Y was $Y_i \sim \text{BB}(N_i,\mu_0,\rho)$ in every center except one, and $Y_M \sim \text{BB}(N_i,\mu_1,\rho)$ in one atypical center, where $\mu_1=\mu_0\pm\delta$.

It was assumed that: (i) all trial centers had the same number $n_i$ of participants; (ii) only one center had an atypical Y distribution. The HBBB method was applied for each dataset generated.

Each scenario was replicated 1000 times to evaluate the sensitivity and specificity of the method in detecting the atypical center. The sensitivity and specificity were calculated by counting the number of true positives (#TP), true negatives (#TN), false positives (#FP), and false negatives (#FN) for each simulation. Sensitivity and specificity were calculated as follows:

$$\text{sensitivity} = \frac{\#\text{TP}}{\#\text{TP}+\#\text{FN}}$$
$$\text{specificity} = \frac{\#\text{TN}}{\#\text{FP}+\#\text{TN}}.$$

### 3.2. Scenarios

The parameters used in the simulations are summarized in Table 1.

In the base case scenario, a set of trials was simulated with the following parameters: total number of trial centers M = 10; number of participants in the atypical center $\text{N}_\text{a}=50$ (therefore: overall number of participants in the study $\text{N}=10\times50=500$ and ratio $\text{N}_\text{a}/\text{N}=10$); proportion of events

**Table 1.** Simulation study: fixed and variable parameters.

| Number of study centers | M | 4–10 |
|---|---|---|
| Sample size in the atypical center | $N_a$ | 10–300 |
| Sample size in non-atypical centers | $N_{na}$ | $=N_a$ |
| Sample size in the overall study | $N$ | $=N_a$*M |
| Binary categorical variable Y | | |
| Proportion* in the atypical center | $\mu_1$ | 0–1.0 |
| Proportion* in non-atypical centers | $\mu_0$ | 0.1–0.5 |
| Difference in proportion | $\delta$ | $\lvert\mu_1-\mu_0\rvert$ |
| Overdispersion (all centers) | $\rho$ | 0–0.1 |

*The proportion is that of the first category of the binary variable. When the proportion is 0.1, 0.2, 0.3, 0.4, and 0.5 in the first category, it is 0.9, 0.8; 0.7; 0.6, and 0.5 in the second category.

in non-atypical centers $\mu_0 = 0.1$ to 0.5; proportion of events in the atypical center $\mu_1 = 0$ to 1.0.

Following this first set of analyses, the number $N_a$ of participants in the atypical center was varied between 10 and 300, keeping the number of centers M at 10 (and therefore the ratio $N_a/N$ at 1/10), and the proportion in the non-atypical center at 0.5. The number M of centers was then varied between 4 and 20, keeping the number of participants $N_a$ in the atypical center at 50 (thus, varying the ratio $N_a/N$ from 1/4 to 1/20) and keeping the proportion $\mu_0$ in non-atypical center at 0.5. Finally, $N_a$ and $N_a/N$ were varied simultaneously.

Centralized statistical monitoring is intended to be carried out several times during the course of a multicenter clinical trial, during recruitment and accumulation of data. This configuration of progressive accumulation of data is taken into account in the simulations through the variation of the number of subjects in the atypical good center and also the variation of the number of centers in the trial. Thus, on the one hand, we successively increase the sample size in the atypical center from 10 subjects per center (corresponding to the situation at the start of the clinical trial) to 300 subjects per center. On the other hand, we are increasing the number of investigative centers in the trial, going from 4 (corresponding to the situation at the start of a clinical trial where all the centers are not yet open) to 20 investigative centers. In all these situations of variations in the sample size parameters of the investigating center and the number of centers in the trial, the performance in terms of sensitivity and specificity of our application is evaluated.

All programs and functions were carried out using the R software (version 4.2.1) and the simulations were carried out on the computing clusters of the Aquitain Intensive Computing Mesocenter (MCIA).

In Step 1, the JAGS software through the R-package jagsUI was used to draw posterior sample of parameters (Plummer 2003; Clark and Altwegg 2019). A vague prior distribution was assumed for both hyperparameters $\mu$ and $\rho$ using a = b=c = d = 1. In the MCMC procedure, 2 chains, 2000 iterations and a burning phase of 1000 iterations was considered. The convergence of chains was checked by using the ratio of inter-chain to intra-chain variances for all sampled parameters, which had to be close to 1.

In step 2, we considered the posterior median as the estimated value of parameters.

In Step 3, we assumed that all centers had the same distribution with a common p. For this, we considered the common pattern for all centers reflecting by $\mu$ and$\rho$ hyperparameters to generate a sample of the probability p. For Step 3 and 4, we considered $L = 2000$.

In this article, we choose to set the threshold at $q = 0.05$ for both simulation studies and application.

## 4. Application Study to Illustrate the Significance of the HBBB Method

The HBBB method was used, with data collected in the Temprano ANRS 12136 trial (TEMPRANO ANRS 12136 Study Group 2015), to assess its potential for use in real multicenter trials.

Temprano was a randomized controlled trial carried out in Abidjan, Côte d'Ivoire, to study the benefits and risks of early antiretroviral treatment and early isoniazid prophylaxis in HIV-infected adults. The study took place from March 2008 to July 2012 at nine health centers. Data were recorded using standardized forms. The enrollment process included a pre-inclusion phase. In this phase, participants who met a number of pre-inclusion criteria were given information about the trial and those who agreed to continue were formally pre-included. They then underwent a number of pre-inclusion tests (mainly biological) and were asked to return a week later, when it was decided whether they could be included in the trial or not. The decision to include them depended on the results of the tests and on the individuals giving their final informed consent, having had time to consider the information they had been given by the trial staff and with the support of their relatives. For pre-included individuals who returned for the inclusion visit and were subsequently not included, the reason for non-inclusion was recorded on a specific "non-inclusion form." The non-inclusion reason variable was multiple choice, one of the choices being "unwillingness to participate." For this application study, non-included individuals whose only reason for non-inclusion was "unwillingness to participate" were considered as "refusing to participate as the only non-inclusion criterion, " and non-included individuals who had at least one other reason for non-inclusion were considered as "any other non-inclusion criteria profile." In this example, the event of interest was therefore refusal to participate. The HBBB method was applied to detect the centers in which the proportion of individuals refusing to participate as the only reason for non-inclusion seemed atypical compared to the others. The method was applied sequentially, and the analysis was carried out each time any of the nine trial centers reached 10, 20, 30, 40, 50, 100, 150, 200, 300, 400, and 500 pre-included participants.

For these applications, we used non-informative priors for $\mu$, and also informative priors for both the average $\mu$ using the average proportion of the previous round, and for the overdispersion$\rho$ using an informative beta distribution.

## 5. Results

### 5.1. Simulation Study

Figure 1 shows the sensitivity and specificity of the HBBB method in detecting an atypical center with three different values for the overdispersion parameter $\rho$ (0, 0.01, 0.1), across a wide distribution range for the variable in the atypical center, and in non-atypical centers. When the delta $\delta = |\mu_1 - \mu_0|$ between the proportion in the atypical center and the non-atypical centers is increased, the sensitivity increases irrespective of the value of $\mu_0$ in the non-atypical centers, but when the value of $\mu_0$ in the non-atypical centers is reduced, it reduces the value of $\delta$ where sensitivity exceeds 90%. When $\rho$ is increased to 0.1, the minimum values of sensitivity increases and specificity decreases. When $\rho$ is 0, specificity exceeds 95% in a wide range of $\mu_0$, $\mu_1$ and $\delta$ values. When $\rho$ is 0.01, specificity exceeds 95% whatever the values of $\mu_0$ and $\delta$.

Figure 2 shows the sensitivity and specificity across a wide range of participant numbers in the atypical center (keeping

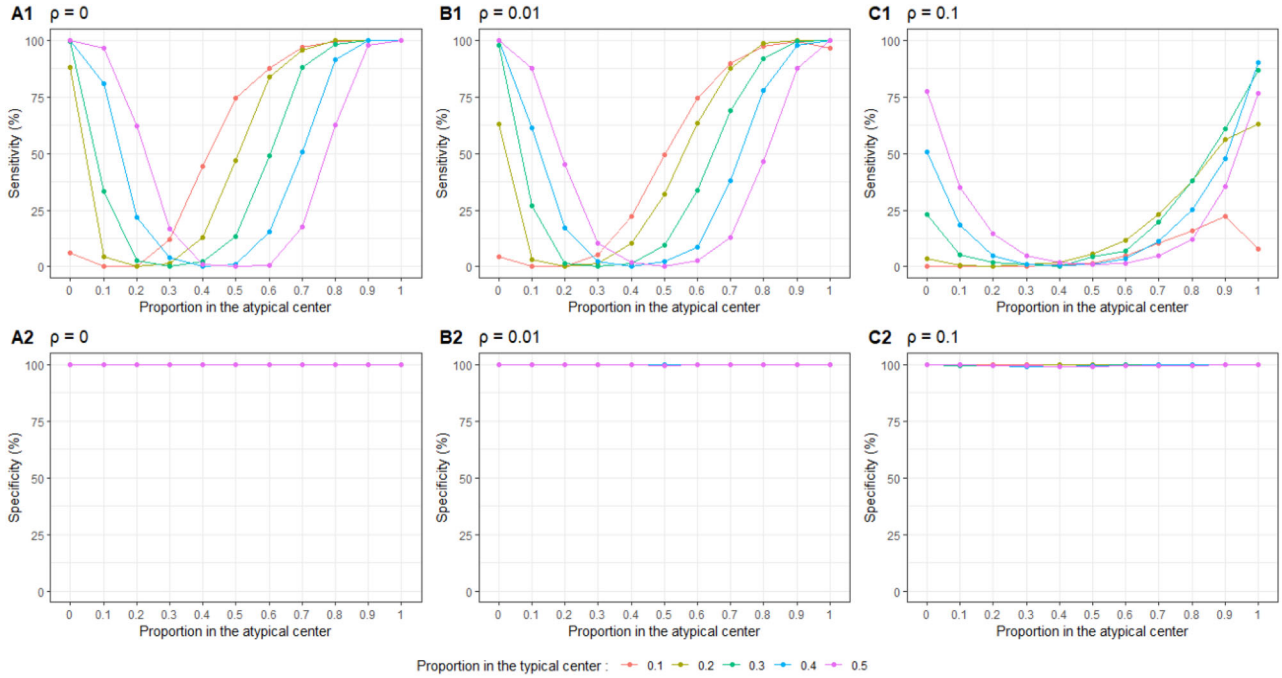


**Figure 1.** Simulation study: Sensitivity and specificity of the HBBB method to detect the atypical center under base case scenario, according to different figures of overdispersion, and different proportions of events in the atypical center ($\mu_1$) and in the other ones ($\mu_0$). This figure explores the sensitivity (A1, B1, C1) and specificity (A2, B2, C2) of the HBBB method to detect an atypical center for the distribution of a categorical variable, across the range of distribution of the categorical variable in the atypical center (0–1.0) (horizontal axis) and in the non-atypical centers (0.1–0.5) (colored lines), according to three overdispersion parameter values (0 in Figures A1 and A2; 0.01 in Figures B1 and B2; 0.1 in Figures C1 and C2). The analysis is performed when the number of participants in the atypical center is 50, in a study in which the total number of participants is ten times higher.

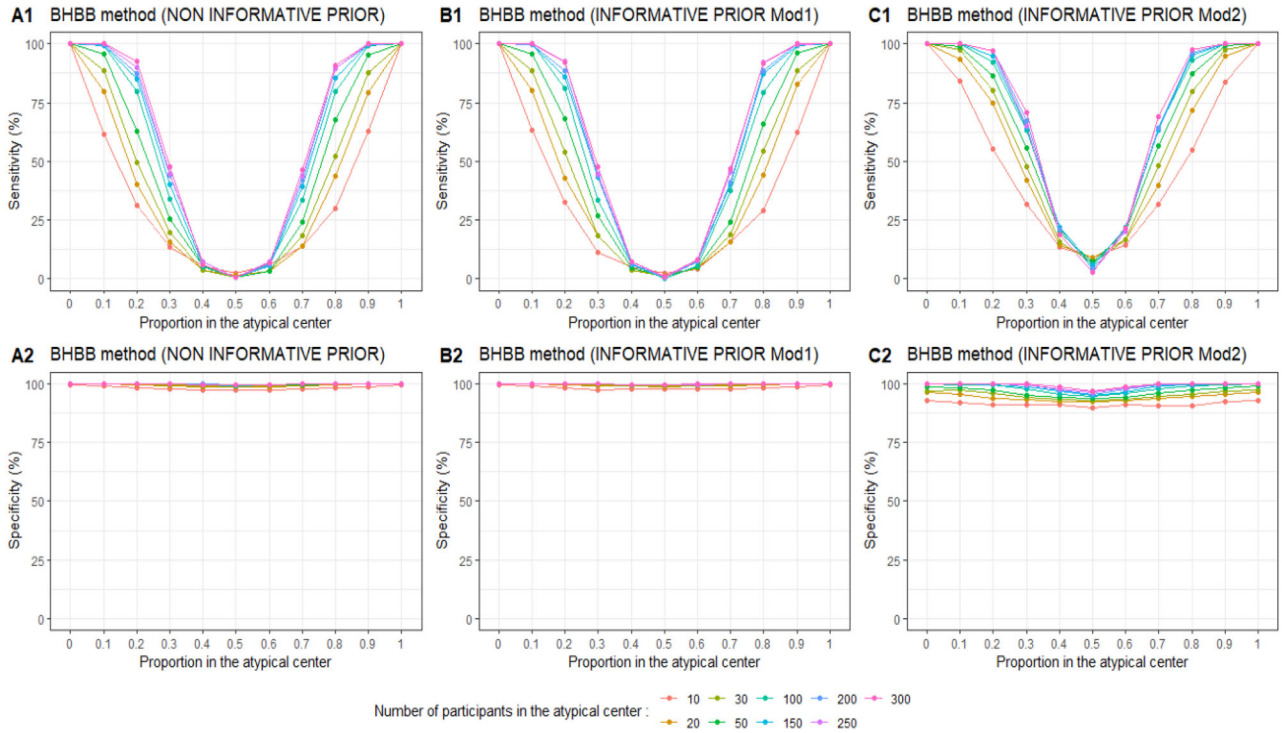


**Figure 2.** Simulation study: Sensitivity and specificity of the HBBB method to detect the atypical center according to the sample size and the proportions of events in the atypical center. This figure explores the sensitivity (A1, B1, C1) and specificity (A2, B2, C2) of the HBBB method to detect an atypical center for the distribution of a categorical variable, across the range of distribution of the categorical variable in the atypical center (0–1.0) (horizontal axis) and according to different number of participants in the atypical center (10–300) (colored curves). The Bayesian model estimation in Figures A1 and A2 uses non-informative priors; while Figures B1 and B2 use an informative prior on the mean (Mod1), and Figures C1 and C2 use an informative prior on the overdispersion (Mod2). The analysis is performed assuming that the total number of participants is 10 times the number of participants in the atypical number. The proportion of events in the non-atypical centers is 0.5.

$N_a/N$ at 1/10) and a wide range of $\mu_1$ proportions in the atypical center (keeping the proportion $\mu_0$ in the non-atypical centers at 0.50), according the possible influences of the choices of priors on the performance of the proposed method.

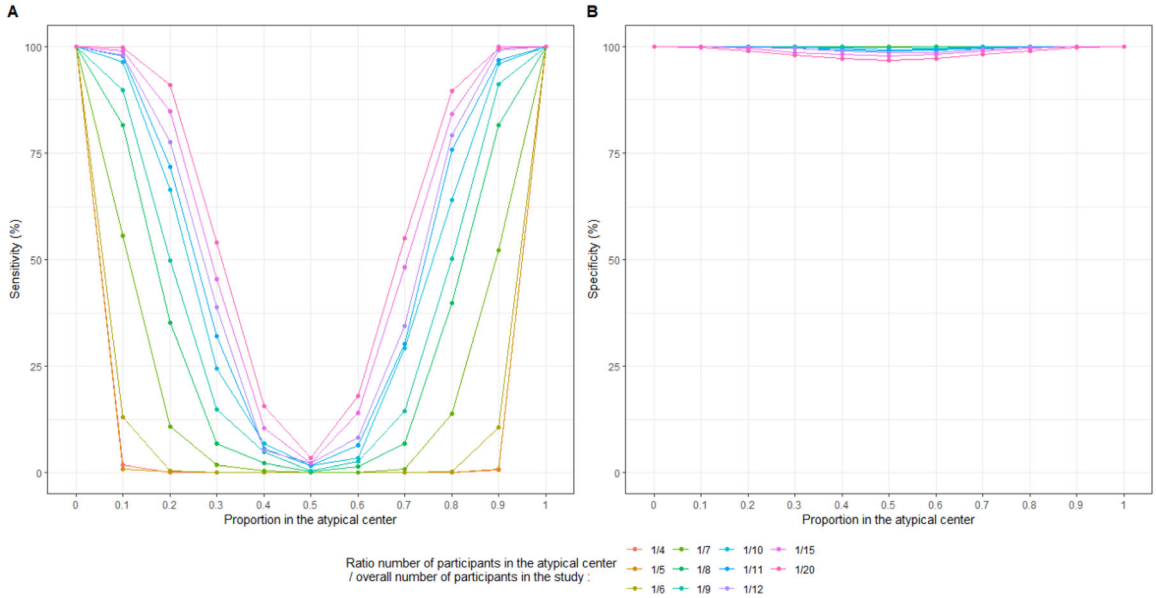


**Figure 3.** Simulation study: Sensitivity and specificity of the HBBB method to detect the atypical center according to the ratio "sample size in the atypical center/sample size in the overall study" and the proportions of events in the atypical center. This figure explores the sensitivity (A) and specificity (B) of the HBBB method to detect that a center is atypical for the distribution of a categorical variable, across all the range of distribution of the categorical variable in the atypical center (0–1.0) (horizontal axis) and according to different ratio $N_a/N$ (1/4–1/20, colored curves). The proportion of events in the non-atypical centers is 0.5. The analysis is performed assuming that the number of participants is 50 in the atypical center (therefore, the total number of participants in the study varies from 200 to 1000).

Despite the simulations carried out using non-informative priors, a sensitivity analysis is performed using informative priors. Thus, the Mod1 model uses an informative prior on the mean ($\mu$) while maintaining a non-informative prior on the overdispersion ($\rho$). This prior leads to putting more weight on the expectation of p around 0.5 while considering no information on the overdispersion of typical proportion p. Then, the Mod2 model uses an informative prior on both the mean ($\mu$) and the overdispersion ($\rho$). In this second informative scenario, we consider the expected proportion of an atypical center to be around 0.5 (prior on mu) with greater certainty (i.e., a tighter distribution of p around 0.5 in Mod2 than in Mod1). For Mod1 and Mod2 models, the considered prior distributions are respectively:

$$\text{Mod1} : \mu \sim \text{Beta}(3,3) \text{ and } \rho \sim \text{Beta}(1,1)$$
$$\text{Mod2} : \mu \sim \text{Beta}(3,3) \text{and } \rho \sim \text{Beta}(0.3,1).$$

Whatever the choice of prior distribution, we have noted that when the number of participants in the atypical center is increased, sensitivity increases in all scenarios, and specificity increases when $\delta$ increases but decreases when $\delta$ decreases.

Figure 3 shows the sensitivity and specificity across a wide range of $N_a/N$ ratios (keeping $N_a$ at 50) and a wide range of $\mu_1$ proportions in the atypical center (keeping the proportion $\mu_0$ in the non-atypical centers at 0.50, and therefore varying $\delta$). When $N_a/N$ is increased, sensitivity in all scenarios decreases, and specificity increases when $\delta$ decreases but decreases when $\delta$ increases.

Figure 4 shows the sensitivity and specificity across a wide range of $N_a$ and $N_a/N$, keeping the $\mu_0$ proportion in the non-atypical centers at 0.50 where $\mu_1$ in the atypical center is 0.1 (and therefore $\delta = 0.4$), 0.2 ($\delta = 0.3$), and 0.3 ($\delta = 0.2$). In this figure, $N_a$ and $\delta$ appear to be more significant determinants of good sensitivity than $N_a/N$. Sensitivities in this figure show two different evolutions: one increasing according to the size of the center and the other decreasing according to the size of the center. For participant ratios varying from 1/8 to 1/20, sensitivity increases from 50% to approximately 100% depending on the increase in the number of subjects per center for $\delta = 0.4$, then from 25% to 100% for $\delta = 0.3$ and 5% to around 80% for $\delta = 0.2$. Furthermore, for ratios from 1/7 to 1/4, we note that the sensitivity decreases with the increase in the number of participants until reaching zero regardless $\delta$. On the other hand, in all scenarios the specificity is conservative and remains above 90%.

### 5.2. Application Study

At the enrollment phase for the Temprano trial, 2962 individuals agreed to be pre-included and 2651 of these returned for the inclusion visits. Of the 2651, 2076 were eventually included and 575 excluded, including 105 for whom the only reason for non-inclusion was unwillingness to participate (TEMPRANO ANRS 12136 Study Group 2015).

In all the sequential analyses described in the methods section, no center is detected as atypical when we use non-informative priors, and two study centers were considered atypical at least once in terms of this variable when we use informative prior. Figure 5 shows the results of the analysis concerning these all centers including the two that were detected (see Appendix). The HBBB method flagged center B as atypical when it reached 50 participants, at which point the overall number of participants in the study was 242, the proportion $\mu_1$ in center B was 22%, and the proportion $\mu_0$ in the overall study was 9.5%. The HBBB method also flagged center C as atypical when it reached 200 participants, at which point the overall

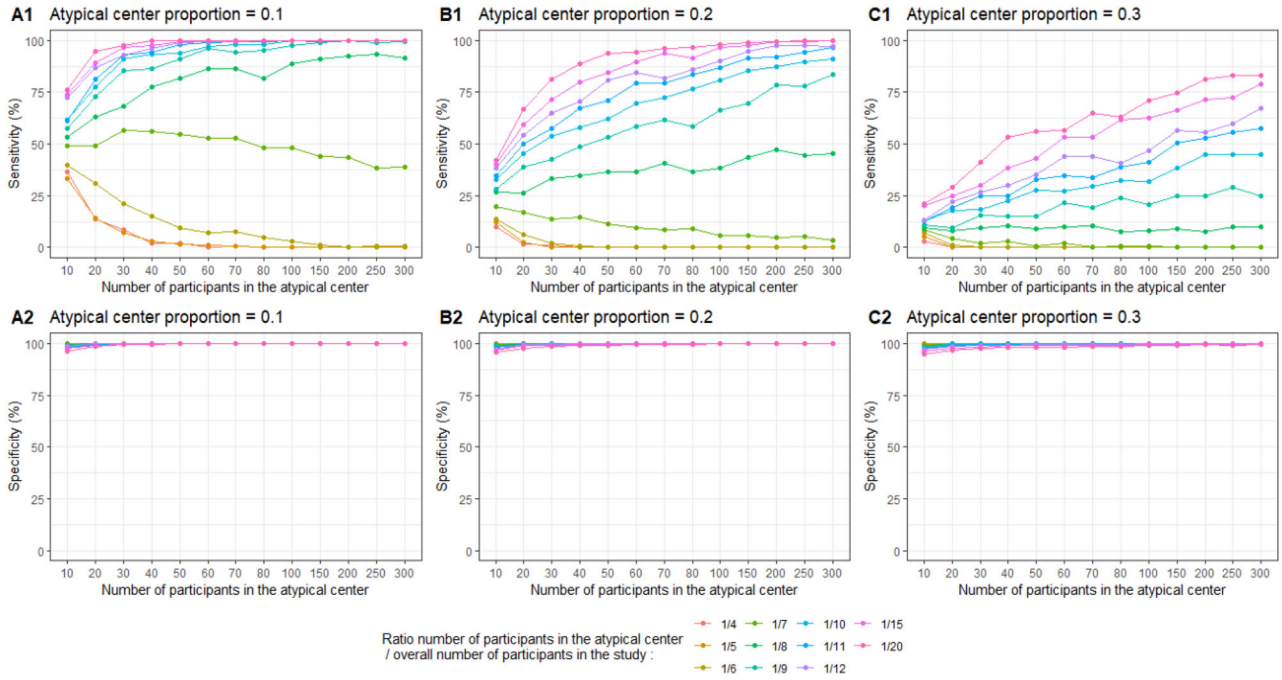


**Figure 4.** Simulation study: Sensitivity and specificity of the HBBB method to detect the atypical center according to the sample size in the atypical center, the ratio "sample size in the atypical center/sample size in the overall study," and the proportions of events in the atypical center. This figure explores the sensitivity (A1, B1, C1) and specificity (A2, B2, C2) of the HBBB method to detect that a center is atypical for the distribution of a categorical variable, according to different number of participants in the atypical center (10–300) (horizontal axis) and different ratio $N_a/N$ (from 1/4 to 1/20) (colored lines). The proportion of events in the non-atypical centers is 0.5. The analysis is performed assuming that the proportion of events in the atypical center is 0.1 (A1 and A2), 0.2 (B1 and B2), and 0.3 (C1 and C2), and therefore the value of $\delta$ is 0.4, 0.3, and 0.2

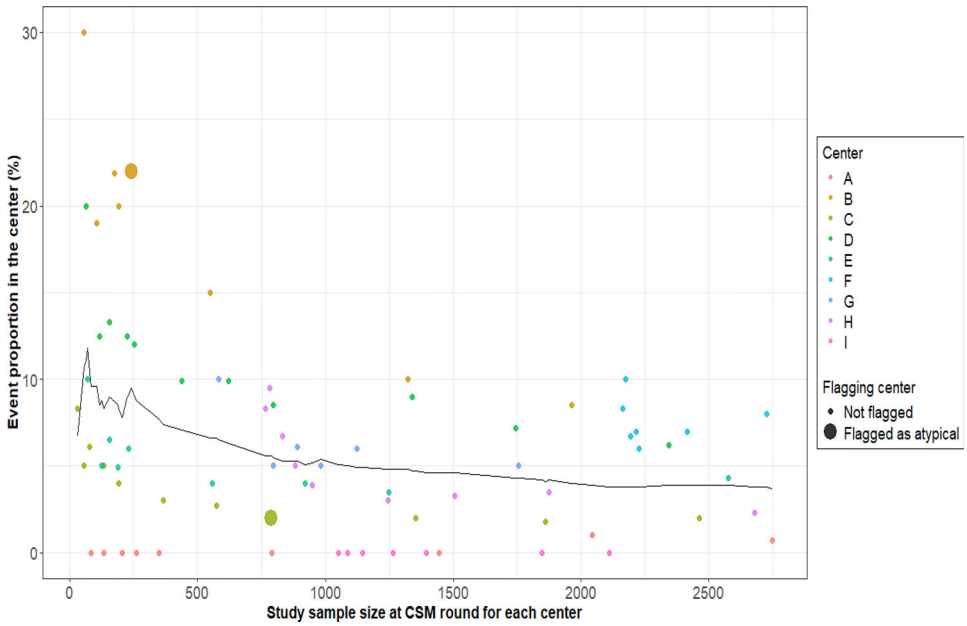


**Figure 5.** Application study: Results of the sequential HBBB analyses in the Temprano trial centers. This figure explores for each center investigated and for each round of CSM the proportion of events in the center (colored dots), that is the proportion of "refusal to participate as the only criterion for non-inclusion" (vertical axis) among the total number of pre-included patients (horizontal axis). The black line in the figure is the event proportion in the overall study. The small dots in the figure indicate cases where the center analyzed in the CSM round is not detected by the CSM method, while the large dots indicate cases where the analyzed center is detected as "atypical".

number of participants in the study was 787, the proportion $\mu_1$ in center C was 2%, and the proportion $\mu_0$ in the overall study was 5.6%. The results of the sequential analyses for the 7 other Temprano trial centers as well as for the two centers detected are also reported in the table in the supplemental appendix.

## 6. Discussion

This article presents a new CSM method for categorical variables to detect atypical centers in multicenter clinical trials. The theoretical conceptualization of the method is based on the principle of mixing binomial models in each center with the probability of

success in mixing data from all centers with a beta distribution (Lee and Sabavala 1987). The HBBB method can be used for categorical data with more than two modalities by regrouping categories to single out the category of interest.

The simulation study suggests that the performance of the HBBB method is similar to the one previously described by Desmet et al. (2017) with regard to the sensitivities and specificities of these two methods. The added value of the HBBB method is that it uses MCMC methods to draw samples from posterior distribution of the parameters of the hierarchical beta-binomial model, using all data already available as a learning sample. This enables the determination of credibility intervals for event occurrences in the form of regions of highest a posteriori density for each center in the multicenter trial. The method is also user-friendly and the results are easy to interpret. It is less complex in theoretical conceptualization because the method is based on beta-binomial modeling and uses a Bayesian approach for parameter estimates, and the software programming is easier as it uses existing packages.

The practical interest of CSM in the field deserves two comments.

First, the sensitivity and specificity of any CSM method should not be considered in the same way as for a diagnostic tool, where the development of close to 100% sensitivity and specificity are generally sought. CSM is not intended to replace other traditional monitoring procedures but is a useful complementary tool that allows other procedures to be optimized by focusing on the potential risks identified. With CSM, poor sensitivity and failure to detect an atypical center can be tolerated while poor specificity cannot, as it generates unnecessary additional control work.

Second, many multicenter trials do not involve hundreds of participants per center. Where they do, it makes less sense to detect potential problems when the sample size is large and/or the trial is nearing completion. It is important, therefore, that the CSM method is capable of detecting atypical centers with sample sizes as low as possible (Niangoran et al. 2023).

Indeed, a CSM method which has very high specificity makes it possible to reduce the number of false positive results to an acceptable level, that is considerably reduce the detection of falsely atypical investigative centers (Power, Fell, and Wright 2013). Thus, it would not generate more field investigation work by data monitors in the centers. However, from a perspective of combining several applications, our HBBB CSM method, although having less good sensitivity, but excellent specificity, could prove very useful in reducing false positives among centers identified as atypical. Also, the judicious choice of the prior distribution for the overdispersion parameter makes it possible to increase sensitivity, while controlling a respectable reduction in specificity. Likewise, increasing the detection threshold of the atypical center (q-parameter) could improve the sensitivity of our CSM application.

The practical application of CSM therefore requires knowledge of how to interpret the results of analyses carried out with various sample sizes and event frequencies, and how to arrive at a compromise between detecting potential problems early and triggering time-consuming control procedures.

In the application study, two atypical centers were convincingly detected, in which $N_a$ was 50 and 200, $\delta$ was 0.12 and 0.04, and $N_a/N$ was 21% and 25%, respectively. This application detected center B as atypical a little earlier in the implementation of the CSM rounds (fifth round), while center C was detected a little later (i.e., in the eighth round). It was noted that the HBBB method that we propose detects both atypical centers with proportions higher than the overall trend, as well as atypical centers with proportions lower than the overall trend. To conclude on the lessons of this application, we note that at relatively small sample sizes, centers with larger differences in proportions compared to that of the whole are more likely to be reported as atypical. On the other hand, we note that for larger sample sizes, even centers presenting proportions not very far from that of the whole are more likely to be reported as atypical by our method.

The CSM study was done retrospectively when the database was closed, but it can reasonably be argued that the use of the HBBB method while the trial was in progress could have led to important practical measures. Prospective studies are nonetheless needed to better assess the benefits and disadvantages of CSM in multicenter trials, for the purposes of comparing all the proposed methods and identifying the strengths and weaknesses of each in different scenarios.

There are several reasons to explain the advantages of the HBBB CSM method that we propose. Among these, there is the advantage of using our method based on Bayesian modeling when few data are available in the multicenter trial and frequentist methods do not make it possible to obtain default results of convergence of the estimates. Likewise, when we have strong a priori knowledge on the distributions of priors in the investigating centers, and which it is possible to integrate into the modeling. Finally, we can consider that the HBBB method that we propose is much simpler in its implementation compared to that proposed by Desmet et al. (2017).

Furthermore, simulation results show that the proposed method controls excellent sensitivity when priors are uninformative. When informative priors are used on mean ($\mu$) and the overdispersion ($\rho$), this improves sensitivity at the expense of specificity. These results are also consistent with those previously obtained by Morita, Thall, and Müller (2010) who considered that well-chosen prior distributions were important to ensure good properties of diagnostic tools, in particular for CSM methods. It is then necessary to take care that the prior information used leads to a reasonable average proportion distribution of typical proportion according to expert knowledge or historical data.

In conclusion, very few CSM methods have been proposed for categorical variables. The performance of this new HBBB method is similar to that previously proposed by Desmet et al. (2017). It is easy to model and shows good specificity in many scenarios where the sample size is limited. We illustrated its potential interest using an historical database. Its practical interest should now be assessed prospectively in future multicenter trials.

## Abbreviations

| | |
|---|---|
| ANRS | French National Agency for Research on AIDS and Viral Hepatitis |
| CSM | Centralized statistical monitoring |
| HBBB | Hierarchical Bayesian beta-binomial |

| | |
|---|---|
| JAGS | Just another Gibbs sampler |
| MCIA | Aquitain intensive computing mesocenter |
| MCMC | Markov chain Monte Carlo |

## Supplementary Materials

Sequential analyses with the HBBB method using the Temprano trial data when each center reached 10, 20, 30, 40, 50, 100, 150, 200 and 300 pre-inclusion.

## Authors' Contributions

Serge Niangoran designed and interpreted the analyses and drafted the first version of the manuscript. Amadou Alioum and Xavier Anglaret contributed to the design and interpretation of the analyses, and substantially revised the manuscript. Antoine Barbieri helped with the Bayesian modeling and theoretical design of the HBBB method. Anani Badjé and Gerard Kouamé contributed in the work on the Temprano database in the application study. Valérie Journot and Olivier Marcy critically revised the article. All authors read and approved the final version of the article.

## Disclosure Statement

All authors declared no competing interests.

## Data Availability Statement

The dataset generated and analyzed during the current study is available from the corresponding author on reasonable request.

## Funding

Serge Niangoran received a scholarship from the French Institut de Recherche pour le Développement (IRD) from January 2020 to December 2022.

## ORCID

Serge Niangoran http://orcid.org/0000-0003-1747-1995

## References

Agency, E. M. (2013), "Reflection Paper on Risk Based Quality Management in Clinical Trials," *Compliance Inspector*, 44, 1–15. [1]

Buyse, M., George, S. L., Evans, S., Geller, N. L., Ranstam, J., Scherer, B., Lesaffre, E., Murray, G., Edler, L., Hutton, J., Colton, T., Lachenbruch, P., and Verma, B. L. (1999), "The Role of Biostatistics in the Prevention, Detection and Treatment of Fraud in Clinical Trials," *Statistics in Medicine*, 18, 3435–3451. DOI:10.1002/(sici)1097-0258(19991230)18:24<3435::aid-sim365>3.0.co;2-o. [1]

Center for Drug Evaluation and Research. (2022), "Oversight of Clinical Investigations—A Risk-Based Approach to Monitoring," *U.S. Food and Drug Administration*, FDA, Available at *https://www.fda.gov/regulatory-information/search-fda-guidance-documents/oversight-clinical-investigations-risk-based-approach-monitoring*. [1]

Chuang-Stein, C. (1993), "An Application of the Beta-Binomial Model to Combine and Monitor Medical Event Rates in Clinical Trials," *Drug Information Journal*, 27, 515–523. DOI:10.1177/009286159302700242. [1]

Clark, A. E., and Altwegg, R. (2019), "Efficient Bayesian Analysis of Occupancy Models with Logit Link Functions," *Ecology and Evolution*, 9, 756–768. DOI:10.1002/ece3.4850. [4]

de Viron, S., Trotta, L., Schumacher, H., Lomp, H.-J., Höppner, S., Young, S., and Buyse, M. (2022), "Detection of Fraud in a Clinical Trial Using Unsupervised Statistical Monitoring," *Therapeutic Innovation & Regulatory Science*, 56, 130–136. DOI:10.1007/s43441-021-00341-5. [1]

Desmet, L., Venet, D., Doffagne, E., Timmermans, C., Burzykowski, T., Legrand, C., and Buyse, M. (2014), "Linear Mixed-Effects Models for Central Statistical Monitoring of Multicenter Clinical Trials," *Statistics in Medicine*, 33, 5265–5279. DOI:10.1002/sim.6294. [1]

Desmet, L., Venet, D., Doffagne, E., Timmermans, C., Legrand, C., Burzykowski, T., and Buyse, M. (2017), "Use of the Beta-Binomial Model for Central Statistical Monitoring of Multicenter Clinical Trials," *Statistics in Biopharmaceutical Research*, 9, 1–11. DOI:10.1080/19466315.2016.1164751. [1,2,8]

Everson, P. J., and Bradlow, E. T. (2002), "Bayesian Inference for the Beta-Binomial Distribution via Polynomial Expansions," *Journal of Computational and Graphical Statistics]*, 11, 202–207. [2]

Hatayama, T., and Yasui, S. (2020), "Bayesian Central Statistical Monitoring Using Finite Mixture Models in Multicenter Clinical Trials," *Contemporary Clinical Trials Communications*, 19, 100566. DOI:10.1016/j.conctc.2020.100566. [1]

Hespanhol, L., Vallio, C. S., Costa, L. M., and Saragiotto, B. T. (2019), "Understanding and Interpreting Confidence and Credible Intervals around Effect Estimates," *Brazilian Journal of Physical Therapy*, 23, 290–301. *https://doi.org/10.1016/j.bjpt.2018.12.006*. [3]

Kleinman, J. C. (1975), "Proportions with Extraneous Variance: two Dependent Samples," *Biometrics*, 31, 737–743. [1]

Kruschke, J. K., and Vanpaemel, W. (2015), "Bayesian Estimation in Hierarchical Models," in *The Oxford Handbook of Computational and Mathematical Psychology*, eds. J. R. Busemeyer, Z. Wang, J. T. Townsend, and A. Eidels, pp. 279–299. Oxford: Oxford University Press. DOI:10.1093/oxfordhb/9780199957996.013.13. [1]

Lee, J. C., and Sabavala, D. J. (1987), "Bayesian Estimation and Prediction for the Beta-Binomial Model," *Journal of Business & Economic Statistics*, 5, 357–367. DOI:10.1080/07350015.1987.10509600. [2,8]

Morita, S., Thall, P. F., and Müller, P. (2010), "Evaluating the Impact of Prior Assumptions in Bayesian Biostatistics," *Statistics in Biosciences*, 2, 1–17. DOI:10.1007/s12561-010-9018-x. [8]

Niangoran, S., Journot, V., Marcy, O., Anglaret, X., and Alioum, A. (2023), "Performance of Four Centralized Statistical Monitoring Methods for Early Detection of an Atypical Center in a Multicenter Study," *Contemporary Clinical Trials Communications*, 34, 101168. DOI:10.1016/j.conctc.2023.101168. [8]

Plummer, M. (2003), "JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling," in 3rd International Workshop on Distributed Statistical Computing (DSC 2003), Vienna, *Austria*, 124. [4]

Pogue, J. M., Devereaux, P. J., Thorlund, K., and Yusuf, S. (2013), "Central Statistical Monitoring: detecting Fraud in Clinical Trials," *Clinical Trials (London, England)*, 10, 225–235. DOI:10.1177/1740774512469312. [1]

Power, M., Fell, G., and Wright, M. (2013), "Principles for High-Quality, High-Value Testing," *Evidence-Based Medicine*, 18, 5–10. DOI:10.1136/eb-2012-100645. [8]

Robert, C. (2007), *The Bayesian Choice, Springer Texts in Statistics*, New York, NY: Springer. DOI:10.1007/0-387-71599-1. [3]

TEMPRANO ANRS 12136 Study Group. (2015), "A Trial of Early Antiretrovirals and Isoniazid Preventive Therapy in Africa," *The New England Journal of Medicine*, 373, 808–822. DOI:10.1056/NEJMoa1507198. [4,6]