Contemporary Clinical Trials Communications 21 (2021) 100709

Contents lists available at ScienceDirect

Contemporary Clinical Trials Communications

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

Research paper

Bayesian survival analysis for early detection of treatment effects in phase 3 clinical trials

Lucie Biard [a,b,c,*], Anne Bergeron [a,b,d], Vincent Lévy [a,e,f], Sylvie Chevret [a,b,c]

[a] INSERM U1153, Team ECSTRRA, Hôpital Saint Louis, 1 avenue Claude Vellefaux, 75010 Paris, France
[b] Université de Paris, Paris, France
[c] AP-HP Hôpital Saint Louis, Service de Biostatistique et Information Médicale, 1 avenue Claude Vellefaux, 75010, Paris, France
[d] AP-HP Hôpital Saint Louis, Service de Pneumologie, 1 avenue Claude Vellefaux, 75010, Paris, France
[e] Université Paris 13, France
[f] AP-HP Hôpital Avicenne, Unité de Recherche Clinique Bobigny, France

ARTICLE INFO

Keywords:
Bayesian inference
Censored data
Historical data
Clinical trial

ABSTRACT

Despite appealing characteristics for the clinical trials setting, Bayesian inference methods remain scarcely used, especially in randomized controlled clinical trials (RCT). This is particularly true when dealing with a survival endpoint, likely due to the additional complexities to model specifications. We propose to use Bayesian inference to estimate the treatment effect in this setting, using a proportional hazards (PH) model for right-censored data. Implementation of such an estimation process is illustrated on two working examples from cancer RCTs, the ALLOZITHRO and the CLL7-SA trials, both originally analyzed using a frequentist approach. In these two different settings, we show that Bayesian sequential analyses can provide early insight on treatment effect in RCTs. Relying on posterior distributions and predictive posterior probabilities, we find that Bayesian sequential analyses of the ALLOZITHRO trial, which was terminated early due to an unanticipated deleterious effect of the intervention on survival, allow quantifying early that the treatment effect was opposite to what was expected. Then, incorporating historical data in the sequential analyses of the CLL7-SA trial would have allowed the treatment effect to be closer to the protocol hypothesis. These *post-hoc* results give grounds to advocate for a wider use of Bayesian approaches in RCTs, including those with right-censored endpoints, as informative decision tools.

1. Introduction

Traditionally, randomized clinical trials (RCT) are designed and analyzed from a frequentist perspective using classical hypothesis testing. However, there is a growing awareness of the usefulness of Bayesian methods in analyzing RCTs, following [1–4]. Indeed, the Bayesian approach possesses a number of practical advantages over the conventional approach that could be used in RCTs: (1) it allows the explicit integration of previous knowledge with new empirical data; (2) it avoids the inevitable misinterpretations of *p*-values [5,6]; (3) it replaces the misleading *p*-value with a summary statistic having a natural and clinically relevant interpretation — the probability that the study hypothesis is true conditioning on the observations; and (4) it is tailored to the learning process: as information becomes available, one updates what one knows, and this gives the Bayesian approach its flexibility and makes it ideal for clinical research. Therefore, it is particularly suited to the sequential analyses of RCTs data.

Actually, Bayesian approach has gained popularity in medical, pharmaceutical, and social science research because it allows researchers to combine prior information with data to model data generating processes; thus, it can incorporate previous knowledge on the likelihood of an event into the interpretation of trial results [3,7].

When planning a RCT, previous data are often available, either in the control group (placebo, standard of care) or in the experimental group from slightly different populations (adult instead of pediatrics, animal studies, *etc.*). For instance, accounting for historical or external data can be part of the trial analysis and it can directly influence the design of the trial itself, in choosing whether to include patients in a control arm or use historical control data, or in estimating the required sample size [8,9].

Nevertheless, Bayesian analyses are still mostly used in early phase trials or for innovative adaptive designs, and often restricted to continuous or binary endpoints [10]. We focused on phase 3 clinical trials with a survival outcome measure, a frequent setting in hemato-oncology, to

* Corresponding author at: AP-HP Hôpital Saint Louis, Service de Biostatistique et Information Médicale, 1 avenue Claude Vellefaux, 75010, Paris, France.
E-mail address: lucie.biard@u-paris.fr (L. Biard).

https://doi.org/10.1016/j.conctc.2021.100709
Received 20 July 2020; Received in revised form 4 December 2020; Accepted 4 January 2021
Available online 9 January 2021

illustrate based on real data, two main interesting uses of the Bayesian approach, as a simple tool for early stopping decisions and in borrowing of external data. Indeed, though these developments are not new, they are poorly used in practice, and Bayesian analyses of phase 3 trials, even recently, mostly use posterior densities of outcomes [11], or only historical controls data [12]. More specifically, the aim of this paper was two-fold: (i) to assess how sequential Bayes analyses may allow early decisions (termination) of the trial; (ii) to assess whether the borrowing of external data (for both the control and the experimental groups) in the analysis would allow optimizing the current trial design. The present work uses data from two example trials, both originally analyzed using a frequentist approach: (i) the ALLOZITHRO trial, which was terminated early due to an unanticipated deleterious effect of the azithromycin over placebo, exemplified by an increased cause-specific hazard of hematological relapse (HR, 1.7; 95%CI, 1.2–2.4; P = 0.002) [13], and (ii) the CLL 2007 SA trial (CLL7-SA) that demonstrated a benefit in progression-free survival of rituximab (RTX) maintenance therapy over standard of care in elderly patients with previously untreated chronic lymphocytic leukemia (HR, 0.55; 95%CI, 0.40–0.75, P = 0.0002) [14].

The paper is organized as follows: Section 1 presents the motivating examples in detail and Section 2 introduces the notations 2.1, the survival models used for the analyses 2.2 and specific methodological aspects of the data (2.3 and 2.4 respectively). Results are reported in Section 3 and some discussion is provided in Section 4.

1.1. ALLOZITHRO trial: Bayesian sequential analyses

The ALLOZITHRO trial (NCT01959100) was a multicenter double-blind placebo-controlled randomized phase 3 trial, which aimed to evaluate the efficacy of azithromycin in the prevention of airflow decline in patients after hematopoietic stem cell transplant (HSCT) [13]. Bronchiolitis obliterans syndrome, which results in airflow decline and respiratory function impairment is a known complication of HSCT, related to chronic graft-*versus*-host disease (GvHD). The trial randomized 465 patients, 231 in the azithromycin arm and 234 in the placebo arm, between February 2014 and August 2015.

The primary endpoint was airflow-decline free (AFD-free) survival at 24 months after randomization as the time from randomization to decline in the respiratory function or death of any cause. Respiratory function was assessed every 6 months, by plethysmography (with pre- and post-bronchodilator spirometry). Observations were censored at the date of last follow-up, in patients without events. It was expected that azithromycin would have a protective effect on the respiratory function and therefore an improved AFD-free survival, resulting in a postulated $\log(HR) = \log(0.64)$ under the alternative, assuming a constant effect over the follow-up (that is, proportional hazards (PH)).

Unexpectedly, the trial was terminated prematurely on December 26, 2016 after the trial Data Safety Monitoring Board (DSMB) alerted on an imbalance in the number of hematological relapses across randomization arms [13]. At that time, enrollment was complete but treatment and follow-up were still on-going for 122 patients.

The statistical analysis plan for the ALLOZITHRO trial relied on frequentist methods, without any planned interim analyses. In the present *post-hoc* re-analysis of the trial, we aimed to illustrate how Bayesian sequential analyses of the AFD-free survival could have informed early on the unexpected adverse outcome of this phase 3 trial.

1.2. CLL7-SA trial: Incorporating historical data

The CLL7-SA trial (NCT00645606) was a multicenter randomized open-label phase 3 trial that was conducted to evaluate the efficacy of 2-year rituximab (RTX) maintenance therapy in elderly patients with previously untreated chronic lymphocytic leukemia (CLL), compared to standard of care (SoC) observation (watchful waiting) [14]. The primary endpoint was progression free survival (PFS). Assuming a 32% relative improvement with RTX in the 36-month PFS (66% *vs.* 50%, $HR = 0.6$), a sample size of 161 events from 542 patients, accounting for 25% drop-out rate during the induction part of the CLL therapy, was computed according to the O'Brien and Fleming design. One interim analysis was planned, after 121 events (75% of 161) had been observed. The inclusion period started on June 10, 2008, and ended on August 14, 2014. Eventually, the interim analysis was performed after 150 events: the efficacy boundary was crossed ($P = 0.0009$) and the trial stopped. Overall, 409 patients were randomized, 202 to RTX maintenance and 207 to standard of care.

At the time of the planning of this trial in 2007, information on RTX maintenance therapy in CLL was scarce. Nevertheless, at the beginning of the inclusion period in December 2010, results of the PRIMA trial (NCT00140582) were published, demonstrating the benefit of 2-year RTX maintenance in patients with follicular lymphoma (FL) receiving a RTX plus chemotherapy regimen as first-line treatment (progression free survival: HR=0.55, 95% CI 0.44–0.68) [15]. Although FL is a different population from CLL, these hematologic malignancies share some characteristics and evolution profiles. They both are indolent B-cell lymphoid malignancies, which progress slowly by acute phases. They both develop similar complications related to the immune system, notably infections. In both cases, there was no curative therapy currently available. Last, in both trials, eligible patients had to be treatment-naive. In that sense, it appeared relevant to consider the FL population from the PRIMA trial, though non perfectly, similar and exchangeable to the CLL population from the CLL7-SA trial. Therefore, results from the PRIMA trial on the effect of RTX in FL patients could provide clinically relevant information on the effect of RTX maintenance in CLL patients. Moreover, we also considered the acceptability of this historical dataset for the combination to the current data [16]. Both trials were large European multicenter randomized phase 3 trials. More precisely, following Pocock's criteria [16]: (i) Treatment regimens were comparable in both the control and treatment groups, between the two trials: both experimental arms consisted in RTX maintenance, with intravenous infusion every 8 weeks for 2 years, with close dosages (375mg/m^2 in the historical trial and 500mg/m^2 in the CLL7-SA trial); both control arms consisted in SoC watchful observation; (ii) The historical study was recent compared to the present trial, with online publication available in December 2010, about 2.5 years after the start of inclusion in the CLL7-SA trial; as described above, the two treatment-naive populations appeared exchangeable, (iii) The main endpoint in our *post-hoc* analysis, was progression-free survival, assessed using international standard criteria in both trials, (iv) Differences between the two trials materialized in patient characteristics: the CLL7-SA trial focused on older patients (above 65 years old), but with good performance status and adequate renal and hepatic function for eligibility, (v) Both studies were conducted in similar settings: they were both large multicenter 1:1 randomized open-label parallel controlled trials sponsored by French collaborative groups specific to the disease of interest: the Groupe d'Étude des Lymphomes de l'Adulte (GELA) for the PRIMA trial and the French Innovative Leukemia Organization (FILO) for CLL7-SA trial, with participating centers belonging to these networks, respectively, and sharing common practice and standards of care, and (vi) There were no further indications to anticipate differing results between the two trials.

In the present *post-hoc* re-analysis, we propose to use Bayesian methods to incorporate information from the PRIMA trial as soon as it was published to the analysis of the CLL7-SA trial, a phase 3 trial with a survival endpoint. In particular, we propose using the power prior approach which allows leveraging the external data [17], accounting for the disease heterogeneity between the current trial (CLL patients) and external information (FL patients).

2. Methods

Both motivating examples used a survival endpoint for the primary efficacy assessment, airflow decline (AFD)-free survival and progression-free survival (PFS) respectively. We assumed a proportional hazards setting to estimate the effect of the treatment, following the original analyses of both trials. Sections 2.1 and 2.2 first present the notations and models used for the Bayesian analyses, while specific aspects are developed thereafter, namely sequential Bayesian analyses in Section 2.3 and incorporation of external individual survival data with the power prior approach in Section 2.4.

2.1. Notations

Let n be the number of observations in the dataset, and let $X_1, \dots, X_n$, denote the right-censored times, *i.e.*, $X_i = min(T_i, C_i)$ and $\delta_i = I(T_i \leq C_i)$ with T_i the failure time and C_i the censoring time of individuals i, $i = 1, \dots, n$. Data are given by $(X_1, \delta_1, Z_1), \dots, (X_n, \delta_n, Z_n)$, where X is the observation time and Z the indicator variable for randomization group ($Z_i = 0$ if the patients is allocated to the control group and $Z = 1$ for the intervention group). We consider the proportional hazards model:

$$T_i|Z_i \sim F(t|Z_i) = 1 - (1 - F_0(t))^{exp(Z_i'\beta)} \quad (1)$$

or, equivalently,

$$H(t|Z_i) = exp(Z_i'\beta)H_0(t)$$

where $F_0(t) = 1 - S_0(t)$ is the baseline distribution function of the right-censored time (either time to airflow decline or death, whatever occurred first, and time to progression or death, for the ALLOZITHRO and CLL7-SA trials, respectively), $H_0(t) = \int_0^t h_0(u)du$ is the cumulative baseline hazard function, and β the log hazard ratio (HR).

2.2. Survival model

Using Bayesian inference, the above model (1) includes two parameters: the treatment effect (β) and the baseline survival (equivalently defined by F_0, S_0, h_0, or H_0). Both must be associated with priors, summarizing, before the trial onset, all external evidence in the effect size and the baseline survival, respectively.

First, the $\log(HR)$ was assumed normally distributed, with mean μ_0 and standard deviation σ_0 [18]:

$$\beta \sim \mathcal{N}(\beta_0, \sigma_0)$$

with prior mean β_0 set at zero, resulting in a reference prior centered on the null hypothesis; such a skeptical prior may be thought of as a handicap that the trial data must overcome in order to provide convincing evidence of benefit [19]. The standard deviation σ_0 can be initially set at a large value to insure weakly informative prior regarding the treatment effect, and thus let the observations drive the analyses (which can be evaluated by comparing results with estimates from likelihood approaches).

Then, for the baseline survival, many modeling options have been proposed in Bayesian survival proportional hazards models, either parametric (exponential or Weibull models) or not (using mixtures of Polya trees or transformed Bernstein polynomials, for instance) [20–23]. We considered two approaches for the prior on the baseline hazard $h_0(t)$: a parametric exponential distribution (constant baseline hazard $h_0(t) = h_0$), or a piecewise constant hazard, sometimes referred to as piecewise exponential model (PEM), to allow more flexibility in the baseline hazard. Of note, more complex piecewise hazards models have been proposed, including spline-based models [23–25].

In the exponential model, the prior for the constant baseline hazard was normal $h_0 \sim \mathcal{N}(0, 20^2)$ [26]. In the PEM, the baseline hazard takes constant values h_k in each time interval I_k defined over the observation period:

$$h_0(t) = \sum_{k=1}^{K} h_k I\{t \in I_k\}$$

where $I_k = (d_{k-1}, d_k], k = 1, \dots, K$, is an interval resulting from the partition in K intervals of the observation period, with $d_0 = 0$ and $d_K = \infty$. The time partition was defined following Murray et al.'s approach [27], with $K = \max\{5, \min(\frac{r}{8}, 20)\}$ intervals, where r is the observed number of events in the current trial dataset, and the left bounds d_{k-1} of the intervals I_k correspond to the $(100\times\frac{k}{K})^{\text{th}}$ percentiles of the event times in the current trial dataset.

Specifically, since we assumed a time-invariant effect of the treatment arm, therefore, it resulted that the count of events Δ_k per interval $I_k = (d_{k-1}, d_k]$, conditionally on treatment, is Poisson-distributed:

$$\Delta_k \sim \text{Poisson}(\lambda_k)$$

with rate $\lambda_k = (d_k - d_{k-1})\exp(\alpha_k + Z'\beta)$ for the whole time interval.

A correlated random-walk process was used as the prior for the baseline hazard: $\alpha_1 \sim \mathcal{N}(\alpha_0, \sigma^2_{\alpha_0}), \alpha_k \sim \mathcal{N}(\alpha_{k-1}, \sigma^2_\alpha)$, with $\alpha_0 = 0$, $\sigma_{\alpha_0} = 10$, and $\sigma_\alpha \sim \text{Uniform}(0.01, 100)$ [9,27]. We chose a random-walk process to allow smoothing of the baseline hazard function over time. Indeed, the vague prior on the first interval initiates the random walk, and then subsequent interval parameters are shrunk toward the previous one, given the random walk process.

The models were estimated with Hamiltonian Monte Carlo (HMC) simulations in Stan on R statistical platform, using the `rstan` and `rstanarm` packages [26,28]. We used 4 chains of each 5000 iterations after warm-up, thinning of 5, yielding 4000 iterations overall to include for the analyses. R code is available on GitHub platform at https://github.com/luciebiard/Bayesian_survival_analysis_phase_3_trials.

2.3. Motivating example 1: Sequential analyses

Given the Bayes approach provides a natural framework for sequential learning, the model described above was fitted to the current data sequentially every six months, monitoring the ALLOZITHRO trial on the basis of the posterior distribution of the $\log(HR)$ of azithromycin on AFD-free survival.

Various normal priors were used for the $\log(HR)$ [18]: (i) reference prior, $\beta_0 = 0$, (ii) enthusiastic or clinical prior $\beta_0 = \log(0.64)$. The latter was chosen consistently with the information available to the investigators at the time of trial planning, and used for sample size calculation [13]. It corresponds to the expected effect of azithromycin on 24-month AFD-survival when the trial was planned [13]. Specifically, the anticipated AFD-free survival at 2 years was 45% in the control group, based on literature reporting the prevalence of AFD in allogeneic HSCT recipients [29], and on the French national registry (Agency for Biomedicine) for the post-transplantation survival estimates (66% after one year and 54% after 2 years at the time of study planning). Moreover a 15% benefit with the experimental treatment on 2-year AFD-free survival was deemed clinically relevant, corresponding to a 0.64 hazard ratio. Furthermore, regarding the variance, it has been shown that, for large balanced trials, the estimated $\log(HR)$ has approximate variance 4 divided by the observed number of events [18,30]. We therefore set $\sigma_0 = 4/10$ to account for very limited pre-existing information on this $\log(HR)$ [1,18,30,31].

2.4. Motivating example 2: Incorporating individual external information

Incorporating historical data to a current analysis relies on the assumption that the different datasets are relevant to the population of interest. Depending on the assumption about the homogeneity across the datasets and populations (identity, exchangeability, bias, *etc.*), different modeling strategies are available [18].

In the present setting of the CLL7-SA trial, we wished to incorporate results from a single external phase 3 trial clinically relevant to the CLL7-SA objective but in a slightly different population. To account for such external data in a discounted manner, we chose to use the power prior approach proposed by [17]. Although the true parameter β, modeling the effect of the treatment RTX, is assumed the same across the datasets, the information borrowed from historical external data is discounted compared to the current data in the estimation model. Briefly, the method is equivalent to shrinking the external sample size by a factor a_0 [8].

Let $D = (n, X, \delta, Z)$ be the current available data, as defined in 2.1, on the time-to-event endpoint according to the randomization arm. Let $D_0 = (n_0, X_0, \delta_0, Z_0)$ be the historical data available, where n_0 denotes the sample size, X_0 the n_0-vector of observed times, δ_0 the n_0-vector of censoring indicator and Z_0 the n_0-vector for the randomization arm indicator.

Let $L(\theta|D)$ be the likelihood for a regression model of the endpoint as a function of Z, with θ the vector of model parameters. In the model described above (Section 2.2), θ is the vector $(\beta, \sigma, \alpha_1, \ldots, \alpha_K, \sigma_\alpha)$. Let $\pi_0(\theta|.)$ denote the joint density of the initial prior distribution for θ. Given the historical data D_0, the power prior, to be used for the current analysis, is given by [32]:

$$\pi(\theta|D_0, a_0) = \frac{L(\theta|D_0)^{a_0}\pi_0(\theta|c_0)}{\int_\Theta L(\theta|D_0)^{a_0}\pi_0(\theta|c_0)d\theta} \quad (2)$$

where c_0 is a specified vector of hyperparameters for the initial prior $\pi_0(\theta|.)$, and $0 \leq a_0 \leq 1$ a scalar parameter that represents the weight of the historical data in estimating the prior for θ. It ranges from 0 (ignoring any previous information from dataset D_0) up to 1 (where the historical data is pooled to the current without leveraging). Conditionally on the value of a_0, the posterior distribution for θ after observing the current data D is given by:

$$\pi(\theta|D, D_0, a_0) = \frac{L(D|\theta)\pi(\theta|D_0, a_0)}{\int_\Theta L(D|\theta)\pi(\theta|D_0, a_0)d\theta} \quad (3)$$

The parameter a_0 can be also considered as an unknown parameter. In that case, a_0 is to be estimated from the datasets (current and historical) and the model includes an hyperprior for a_0, $\pi(a_0|\gamma_0)$, with hyperparameter γ_0. As explained by Ibrahim & Chen [32], it is reasonable to set a beta prior for a_0, such as $0 \leq a_0 \leq 1$, although other choices such as truncated gamma or normal prior distributions are possible. When considering a beta prior for a_0, they argued that it is easier to elicit a mean μ_{a_0} and standard deviation σ_{a_0} for a_0 from physicians, rather than directly setting the hyperparameters vector for the beta distribution $\gamma_0 = (p_0, q_0)$; then, hyperparameters can be derived by back substitution using the equations for a beta distribution, as follows [32]:

$$\mu_{a_0} = \frac{p_0}{p_0 + q_0}; \quad \sigma^2_{a_0} = \frac{\mu_{a_0}(1 - \mu_{a_0})}{p_0 + q_0 + 1}.$$

Nevertheless, one should note that, in the case of a random a_0, we obtain a joint posterior distribution, $\pi(\theta, a_0|D_0)$, which must include the normalizing constant [33]:

$$C(a_0) = 1/\int L(D_0|\theta)^{a_0}\pi(\theta|c_0)\mathrm{d}\theta.$$

In the present study, we chose to set a_0 fixed instead of random, given the single historical dataset used to enrich the CLL7-SA analysis, thus the rather limited information to estimate between-trial information via a_0. Nevertheless, sensitivity analyses with several values for a_0 (1, 0.75, 0.5, 0.25, 0) were performed, to assess the robustness of the results to the choice of a_0, that is, to the influence of the historical data on the estimation.

In this example, we used a weakly informative reference prior distribution $\pi_0(\beta)$ for the historical PRIMA data $\mathcal{N}(\beta_0 = 0, \sigma_0 = 10)$.

3. Results

3.1. ALLOZITHRO example

We performed *post-hoc* sequential analyses on the AFD-free survival probability estimated on all available data truncated every 6 months starting from the beginning of enrollment in the trial. Given the time-to-event endpoint, we applied non-informative administrative right-censoring at the cut-off sequential dates. Table 1 reports the timepoints and the corresponding available data of each resulting interim analysis. Inclusions were completed by 15 August 2015. In December 2016, when the intervention was terminated early, 218 patients had been randomized less than 24 months before (primary endpoint observation window), of whom 96 had already experienced a primary event, and 122 had not.

At the first (Aug, 2014) and second (Feb, 2015) interim timepoints, few data on the primary endpoint were available: only 6 and 43 patients had experienced an event at these cut-off dates, respectively (Table 1). For the purpose of the present re-analyses, due to the small number of events, the parametric exponential model was used for these timepoints. The 4 remaining analyses used the more flexible PEM model as described in Section 2.2.

Posterior estimates of the $\log(HR)$ with reference and enthusiastic priors are reported in Table 2. Since the second interim analysis, in February 2015, the mean and median posterior $\log(HR)$ estimates were consistently above 0, whatever the prior. Moreover, the 10^{th} percentile of the posterior $\log(HR)$ distribution was consistently above 0, starting from the fourth analysis (February 2016), pointing toward a probable increased risk of event in patients treated with azithromycin. In other words, starting from this date, based on the $\log(HR)$ credibility intervals, there was a probability lower than 5% that azithromycin was beneficial in terms of AFD-free survival (with the lower bound of the $\log(HR)$ 95% credibility interval close to 0); conversely, if we consider the similitude with a one-sided hypothesis, there was a probability close to 95% that azithromycin was harmful.

3.2. CLL7-SA example

We examined how published results of the PRIMA trial could have been used, right away after its publication, on December, 2010, to provide information on the effect of RTX in the CLL7-SA trial. More specifically, we aim to illustrate how this information may have been incorporated in an interim Bayesian analysis. At that time, 216 patients had been randomized, with an actual median follow-up 19.1 months, and 20 and 11 observed failures in the SoC and RTX arms, respectively. Kaplan–Meier estimates of the 24-month PFS were 75% (95%CI 65;86) for the SoC group, and 87% (78;96) for the RTX group.

Since we did not have access to the original individual data from the PRIMA trial, we used reconstructed observations from the initial publication of the PRIMA trial. Specifically, based on the published Kaplan–Meier curves and numbers of patients at risk, using DigitizeIt software and iterative numerical methods solving the inverted Kaplan–Meier equations, we obtained a reconstructed dataset mimicking the PRIMA trial results [34]. Briefly, for each time interval reported on the publication, the algorithm combines published numbers of at-risk patients, Kaplan–Meier curve coordinates, and iterative calculations using the Kaplan–Meier estimator. We refer the reader to the original publication for a detailed presentation of this algorithm and the corresponding R code [34]. The reconstructed data was consistent with the published results, yielding a HR=0.54 (95%CI 0.44;0.68) *versus* HR=0.55 (95%CI 0.44;0.68) in the original publication [15].

Fig. 1 presents the Kaplan–Meier estimates of the December 2010 interim CLL7-SA data, with the reconstructed PRIMA estimates superimposed, for progression-free survival in patients with RTX maintenance or SoC observation.

Table 1
ALLOZITHRO example: Sequential timepoints and corresponding samples.

Interim	Date cut-off	Placebo:Azithromycin		
		No. of inclusions	No. of completed follow-up	No. of events
1	August 13, 2014	70:65	3:3	3:3
2	February 13, 2015	149:143	20:23	20:23
3	August 13, 2015	231:234	53:57	53:57
4	February 13, 2016	231:234	70:94	70:94
5	August 13, 2016	231:234	113:138	90:118
6	February 13, 2017	231:234	154:162	110:131

Table 2
ALLOZITHRO example: Sequential posterior estimates of the $\log(HR)$ on AFD-free survival for azithromycin compared to placebo, with either the reference prior (Ref.): $\beta \sim \mathcal{N}(0, 4/10)$, or the enthusiastic clinical prior (Enthu.): $\beta \sim \mathcal{N}(\log(0.64), 4/10)$.

Interim	Date	Prior	Mean $\log(HR)$	Median $\log(HR)$	95% CrI	10th percentile	$Pr(HR > 1)$
1	Aug, 2014	Ref.	−0.019	−0.027	−0.974 ; 0.946	−0.642	0.483
		Enthu	−0.285	−0.288	−1.290; 0.715	−0.938	0.288
2	Feb, 2015	Ref.	0.138	0.141	−0.392 ; 0.679	−0.219	0.688
		Enthu	0.046	0.048	−0.504 ; 0.572	−0.302	0.560
3	Aug, 2015	Ref.	0.118	0.117	−0.243; 0.482	−0.121	0.726
		Enthu	0.077	0.074	−0.278; 0.436	−0.152	0.662
4	Feb, 2016	Ref.	0.317	0.317	0.025; 0.615	0.124	0.982
		Enthu	0.290	0.290	−0.002; 0.581	0.092	0.974
5	Aug, 2016	Ref.	0.318	0.319	0.047; 0.591	0.140	0.989
		Enthu	0.292	0.292	0.018; 0.560	0.118	0.983
6	Feb, 2017	Ref.	0.229	0.228	−0.022; 0.475	0.069	0.964
		Enthu	0.207	0.207	−0.039; 0.460	0.042	0.948

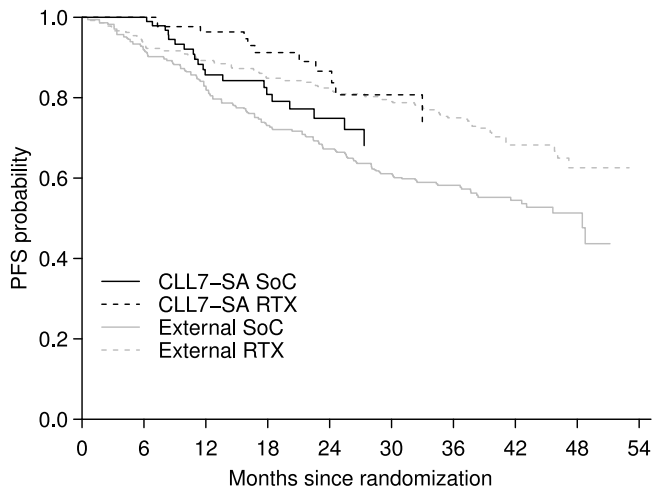

Fig. 1. CLL7-SA example: Kaplan–Meier estimates of progression free survival with interim CLL7-SA data censored on 31 December 2010 (observation standard of care SoC: solid black line; RTX maintenance: dashed black line) and reconstructed data from the PRIMA trial published in January 2011 (observation standard of care SoC: solid gray line; RTX maintenance: dashed gray line).

Results are reported in Table 3. Combined with a weakly informative prior on the effect of RTX (with $a_0 = 0$ and $\mathcal{N}(\beta_0 = 0, \sigma_0 = 10)$ as aggregate weakly informative prior on the treatment coefficient, see Section 2.4), available data in January 2011 pointed toward a reduced hazard of event with RTX maintenance compared to SoC watchful waiting, with mean $\log(HR) < 0$ and 90th percentile < 0. Incorporating information from the PRIMA trial, enriching the prior for analysis of the current data, resulted in a posterior probability of the $\log(HR)$ being lower than 0 greater than 97.5%, even when largely down-weighting the external data ($a_0 = 0.1$). The upper bound of the 95% $\log(HR)$ credibility interval remained consistently lower than 0 as soon as we incorporated some degree of historical data, corresponding to a posterior probability that RTX was harmful lower than 2.5%. Considering an equivalence region, for instance $-0.1 < \log(HR) < 0.1$, there was a posterior probability lower than 1% that RTX was equivalent to SoC, with $a_0 \geq 0.1$ (7% with $a_0 = 0$). Furthermore, with power parameter $a_0 = 0.5$, there was a posterior 70% probability that the HR was lower than 0.6 which was the desired efficacy level used in the frequentist trial planning.

4. Discussion

In this article, we presented how Bayesian inference may inform on time-to-event endpoints in phase 3 clinical trials. Compared to the frequentist approach, which remains widely used in large confirmatory trials, Bayesian methods have been advocated for the straightforward and intuitive interpretation of results, in the form of the posterior estimates of the treatment effect, and their flexibility, in particular for the design of complex trials, such as adaptive designs and borrowing of external data [2–4,7]. Nevertheless, their use requires precautions as issued in guidelines and guidance to prevent misuse and erroneous conclusions by regulatory agencies [35].

We used two different RCTs to illustrate several of the advantages of the Bayes approaches in specific but non rare settings. First, the ALLOZITHRO trial illustrates the contribution of Bayesian inference for sequential analyses of a right-censored endpoint. Moreover, it provides a flexible framework to detect early departures of the treatment effect from the expected direction. Using different prior distributions for the $\log(HR)$, we found consistent results indicating a deleterious outcome for treated patients, more than 6 months before the trial was stopped. Otherwise, the CLL7-SA trial illustrated how the power prior method, borrowing external information, can increase the information on the right-censored endpoint. In both examples, implementing decision rules based on these Bayesian analyses might have had a direct clinical benefit for patients by shortening the trial duration: by discontinuing treatment earlier and preventing prolonged exposure, and by concluding to efficacy earlier and accelerating access to the drug, respectively.

We chose proportional hazards (PH) models, which were used originally in both trials and are the most common in large clinical trials with survival endpoints. The piecewise constant baseline hazard allows more flexibility than the constant hazard exponential model, though it could fail to adequately fit the data in sparse settings. Indeed, the model relies on a partition of the time scale into intervals based on the distribution of failure times. At early interim analyses, there

Table 3
CLL example: posterior distribution of the effect of RTX on PFS ($\log(HR)$) compared to SoC at December 2010 interim analysis of CLL7-SA trial, with reconstructed data from the PRIMA trial results as historical prior information, using the power prior approach with fixed power parameter, ranging from 0 (equivalent to non including historical data) to 1 (equivalent to pooling historical to current data).

Power prior	Mean $\log(HR)$	Median $\log(HR)$	95% CrI	90th percentile	$Pr(HR < 0.6)$
$a_0 = 0$	−0.562	−0.562	−1.322 ; 0.151	−0.094	0.546
$a_0 = 0.10$	−0.586	−0.586	−0.996 ; −0.191	−0.320	0.647
$a_0 = 0.25$	−0.590	−0.590	−0.979 ; −0.221	−0.344	0.665
$a_0 = 0.50$	−0.604	−0.604	−0.883 ; −0.320	−0.421	0.741
$a_0 = 0.75$	−0.604	−0.604	−0.848 ; −0.362	−0.441	0.774
$a_0 = 1$	−0.608	−0.608	−0.815 ; −0.405	−0.472	0.825

might be a limited number of observed failures, which may result in problematic estimations. For instance in our examples, there were 6 and 43 events in the first two interim analyses in the ALLOZITHRO example. Furthermore, in the CLL7-SA trial they were 31 observed events, and the incorporation of external individual data did not allow to reach convergence. Using the partition rule proposed by Murray et al. [27], the PEM model for these analyses could not be estimated without convergence issues, using the Hamiltonian Monte Carlo (HMC) algorithm; this explains why we used the exponential model with a one-parameter baseline hazard. The model could nevertheless be estimated with another MCMC sampler, namely a random-walk Metropolis algorithm (using `MCMCpack` R library [36]), yielding consistent results with those of the exponential model. This limitation could be also tempered by planning the interim analyses according to the expected rate of events. Otherwise, whether reparameterization of the model or other time partition rules could allow convergence of the PEM model in sparse settings, requires further investigations.

Note that the PEM random-walk model can be implemented for estimation in any Bayesian software, on various platforms, as mentioned above: in the present work, we used Stan via the `rstan` package on R platform. Several tools are notably available for convergence diagnosis of Stan HMC estimated models (*e.g.* package `shinystan` [37]). In our setting of time-to-event endpoints, we specifically implemented predicted survival for model checking (see Supplementary material).

More complex survival models, adapted to specific situations, are available and could be applied for these Bayesian analyses allowing non proportional hazards (PH) and time-dependent treatment effect, as well as interval censoring [38–40]. Of note, in the context of non PH issues, [41] proposed to combine current and external data using Bayesian methods, to infer on restricted mean survival.

In the second example, we illustrated how external information can be borrowed to enrich the current data. To that aim, various Bayesian approaches have been proposed which mainly differ in the assumptions about the relevance and exchangeability of the external data with the current trial [18,42,43]. We chose the power prior approach to down-weight the reconstructed data from the PRIMA trial published results, to account for the similar but different disease population (chronic lymphocytic leukemia *versus* follicular lymphoma). Alternatively, we could have used an informative Gaussian prior for the $\log(HR)$ defined based on the PRIMA results, rather than the reconstructed individual data, and discounted this external information by increasing the prior variance on the $\log(HR)$ [3]. In the case several external sources are available, more complex models with hyperparameterization for the between-source variability, can be considered: Bayesian hierarchical modeling and meta-analytical approaches, power priors with random power parameter a_0 [8,9,41,44].

We presented *post-hoc* analyses of two trials to advocate the use of Bayesian methods in phase 3 trials with survival endpoints. Bayesian posterior estimates are particularly adapted for decision rules. Similarly, posterior predictive estimates can also be used to base decision rules on predictions of interest [45]. Formal assessment of the resulting operating characteristics, similarly to the sample size calculation in the frequentist setting, may appear necessary to implement these tools in practice, to comply with the regulatory agencies requirements; guidance to prevent misuse and erroneous conclusions have been issued to this aim [35,46,47]. Indeed, defining rules for efficacy based on conciliatory thresholds, such as the posterior probability of the HR being greater than 1 for instance, often result in unacceptable type I error rates. Last, using predictive probabilities, methods have been developed to estimate the probability of success of a trial at the planning stage, but also during the trial, in a sequential manner using both current and external information [31,48], that could apply in this setting.

In summary, we exemplified throughout two main examples, the informativeness of Bayesian methods in sequential analysis of RCTs with right-censored endpoints. We showed that the Bayesian approach can be applied to proportional hazards survival models with estimation tools available on software platforms and should not be restricted to binary endpoints. Furthermore, we illustrated two aspects of Bayesian methods for phase 3 clinical trials, namely flexible sequential analyses and incorporation of external or historical data. Overall, Bayesian methods provide straightforward interpretation of results, accounting for uncertainty, and allows borrowing information, summarizing all the evidence available at the current time.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

LB was supported by the TRT_cSVD project, France (*From Target Identification to Next Generation Therapies for Cerebral Small Vessel Diseases*, Pr Hugues Chabriat & Dr Anne Joutel, *RHU–Agence Nationale pour la Recherche*, grant number: ANR-16-RHUS-004).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.conctc.2021.100709.

References

[1] D. Spiegelhalter, L. Freedman, M. Parmar, Bayesian approaches to randomized trials, J. Royal Stat. Soc. Appl. Stat. 157 (3) (1994) 357–387.
[2] D. Wijeysundera, P. Austin, J. Hux, W. Beattie, A. Laupacis, Bayesian statistical inference enhances the interpretation of contemporary randomized controlled trials, J. Clin. Epidemiol. 62 (2009) 13–21.
[3] E. Goligher, G. Tomlinson, D. Hajage, et al., Extracorporeal membrane oxygenation for severe acute respiratory distress syndrome and posterior probability of mortality benefit in a post hoc Bayesian analysis of a randomized clinial trial, JAMA 320 (21) (2018) 2251–2259.
[4] S. Ruberg, F. Harrell, M. Gamalo-Siebers, J. LaVange, K. Price, C. Peck, Inference and decision making for 21st-century drug development and approval, Amer. Statist. 73 (2019) 319–327.
[5] V. Amrhein, S. Greenland, B. McShane, Retire statistical significance, Nature 567 (2019) 305–307.
[6] R.L. Wasserstein, A.L. Schirm, A. LN, Moving to a world beyond p < 0.05, Amer. Statist. 73 (2019) 1–9.

[7] J. Bittl, Y. He, Bayesian analysis: A practical approach to interpret clinical trials and create clinical practice guidelines, Circ. Cardiovasc. Qual. Outcomes 10 (2017) e003563.

[8] B. Neuenschwander, G. Capkun-Niggli, M. Branson, D.J. Spiegelhalter, Summarizing historical information on controls in clinical trials, Clin. Trials 7 (1) (2010) 5–18.

[9] J. van Rosmalen, D. Dejardin, Y. van Norden, B. Löwenberg, E. Lesaffre, Including historical data in the analysis of clinical trials: Is it worth the effort? Stat. Methods Med. Res. 27 (10) (2018) 3167–3182.

[10] G. Yin, C.K. Lam, H. Shi, Bayesian randomized clinical trials: From fixed to adaptive design, Contemp. Clin. Trials 59 (2017) 77–86.

[11] M. James Brophy, Bayesian interpretation of the EXCEL trial and other randomized clinical trials of left main coronary artery revascularization, JAMA Internal Med. 180 (7) (2020) 986–992.

[12] A. Bertsche, F. Fleischer, J. Beyersmann, G. Nehmiz, Bayesian Phase II optimization for time-to-event data based on historical information, Stat. Methods Med. Res. 28 (4) (2019) 1272–1289.

[13] A. Bergeron, S. Chevret, A. Granata, et al., Effect of azithromycin on airflow decline free survival after allogeneic hematopoietic stem cell transplant: The ALLOZITHRO randomized clinical trial, JAMA 318 (6) (2017) 557–566.

[14] C. Dartigeas, E. Van Den Neste, J. Léger, et al., Rituximab maintenance versus observation following abbreviated induction with chemoimmunotherapy in elderly patients with previously untreated chronic lymphocytic leukaemia (CLL 2007 SA): an open-label, randomised phase 3 study, Lancet Haematol. 5 (2) (2018) e82–e94.

[15] G. Salles, J.F. Seymour, F. Offner, et al., Rituximab maintenance for 2 years in patients with high tumour burden follicular lymphoma responding to rituximab plus chemotherapy (PRIMA): a phase 3, randomised controlled trial, Lancet 377 (9759) (2011) 42–51.

[16] S.J. Pocock, The combination of randomized and historical controls in clinical trials, J. Chronic Dis. 29 (3) (1976) 175–188.

[17] J.G. Ibrahim, M.H. Chen, Y. Gwon, F. Chen, The power prior: theory and applications, Stat. Med. 34 (28) (2015) 3724–3749.

[18] D.J. Spiegelhalter, K.R. Abrams, J.P. Myles, Bayesian Approaches To Clinical Trials and Health-Care Evaluation, John Wiley & Sons, 2004.

[19] D. Spiegelhalter, J. Myles, D. Jones, K. Abrams, An introduction to Bayesian methods in health technology assessment, Br. Med. J. 319 (1999) 508–512.

[20] I.K. Omurlu, M. Ture, K. Ozdamar, Bayesian analysis of parametric survival models: A computer simulation study based informative priors, J. Stat. Manag. Syst. 18 (5) (2015) 405–423.

[21] H. Zhou, T. Hanson, J. Zhang, `spBayesSurv`: Fitting Bayesian spatial survival models using R, J. Stat. Softw. 92 (9) (2020) 1–33.

[22] M. De Iorio, W. Johnson, P. Muller, G. Rosner, Bayesian nonparametric non-proportional hazards survival modelling, Biometrics 65 (2009) 762–771.

[23] L. Biard, A. Bergeron, S. Chevret, Bayesian models for survival data of clinical trials: Comparison of implementations using R software, 2019, arXiv:1908.06687v3.

[24] T.A. Murray, B.P. Hobbs, D.J. Sargent, B.P. Carlin, Flexible Bayesian survival modeling with semiparametric time-dependent and shape-restricted covariate effects, Bayesian Anal. 11 (2) (2016) 381–402.

[25] S.L. Brilleman, E.M. Elci, J. Buros Novik, R. Wolfe, Bayesian survival analysis using the `rstanarm` R package, 2020, arXiv:2002.09633.

[26] B. Goodrich, J. Gabry, I. Ali, S. Brilleman, `rstanarm`: Bayesian applied regression modeling via Stan, in: R Package Development Version, Survival Branch, 2019.

[27] T.A. Murray, B.P. Hobbs, T.C. Lystig, B.P. Carlin, Semiparametric Bayesian commensurate survival model for post-market medical device surveillance with non-exchangeable historical data, Biometrics 70 (1) (2014) 185–191.

[28] B. Carpenter, A. Gelman, M.D. Hoffman, et al., Stan: A probabilistic programming language, J. Stat. Softw. 76 (1) (2017) 1–32.

[29] J.W. Chien, P.J. Martin, T.A. Gooley, M.E. Flowers, S.R. Heckbert, W.G. Nichols, J.G. Clark, Airflow obstruction after myeloablative allogeneic hematopoietic stem cell transplantation, Am. J. Respir. Crit. Care Med. 168 (2) (2003) 208–214.

[30] A.A. Tsiatis, A the asymptotic joint distribution of the efficient scores test for the proportional hazards model calculated over time, Biometrika 68 (1) (1981) 311–315.

[31] K. Rufibach, P. Jordan, M. Abt, Sequentially updating the likelihood of success of a phase 3 pivotal time-to-event trial based on interim analyses or external information, J. Biopharm. Statist. 26 (2016) 191–201.

[32] J.G. Ibrahim, M.H. Chen, Power prior distributions for regression models, Stat. Sci. 15 (1) (2000) 46–60.

[33] B. Neuenschwander, M. Branson, D.J. Spiegelhalter, A note on the power prior, Stat. Med. 28 (28) (2009) 3562–3566.

[34] P. Guyot, A. Ades, M.J. Ouwens, N.J. Welton, Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan–Meier survival curves, BMC Med. Res. Methodol. 12 (1) (2012) 9.

[35] FDA, Interacting with the FDA on Complex Innovative Trial Design for Drugs and Biological Products, Draft Guidance for Industry, 2019.

[36] D. Andrew, A.D. Martin, K.M. Quinn, J.H. Park, `MCMCpack`: Markov chain Monte Carlo in R, J. Stat. Softw. 42 (9) (2011) 22.

[37] J. Gabry, D. Simpson, A. Vehtari, M. Betancourt, A. Gelman, Visualization in Bayesian workflow, J. Roy. Statist. Soc. Ser. A 182 (2) (2019) 389–402.

[38] X. Wang, M. Chen, J. Yan, Bayesian dynamic regression models for interval censored survival data with application to children dental health, Lifetime Data Analysis 19 (3) (2013) 297–316.

[39] C. Anderson-Bergman, `icenReg`: Regression models for interval censored data in R, J. Stat. Softw. 81 (12) (2017) 1–23.

[40] C. Pan, B. Cai, L. Wang, X. Lin, `ICBayes`: Bayesian Semiparametric Models for Interval-Censored Data. R Package Version 1.1, 2017.

[41] R.J. Klement, P.S. Bandyopadhyay, C.E. Champ, H. Walach, Application of Bayesian evidence synthesis to modelling the effect of ketogenic therapy on survival of high grade glioma patients, Theor. Biol. Med. Model. 15 (1) (2018) 12.

[42] B.P. Hobbs, B.P. Carlin, S.J. Mandrekar, D.J. Sargent, Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials, Biometrics 67 (3) (2011) 1047–1056.

[43] K. Viele, S. Berry, B. Neuenschwander, et al., Use of historical control data for assessing treatment effects in clinical trials, Pharm. Stat. 13 (1) (2014) 41–54.

[44] T.P. Debray, J.A. Damen, R.D. Riley, et al., A framework for meta-analysis of prediction model studies with binary and time-to-event outcomes, Stat. Methods Med. Res. 28 (2019) 2768–2786.

[45] L.Y. Inoue, P.F. Thall, D.A. Berry, Seamlessly expanding a randomized phase II trial to phase III, Biometrics 58 (4) (2002) 823–831.

[46] L. Freedman, D. Spiegelhalter, M.K. Parmar, The what why and how of Bayesian clinical trials monitoring, Stat. Med. 13 (1994) 1371–1383.

[47] H. Shi, G. Yin, Control of type I error rates in Bayesian sequential designs, Bayesian Anal. 14 (2) (2019) 399–425.

[48] Y. Wang, H. Fu, P. Kulkarni, C. Kaiser, Evaluating and utilizing probability of study success in clinical development, Clin. Trials 10 (2013) 407–413.