



HAL
open science

Approche Bayésienne de la survie dans les essais cliniques pour les cancers rares

Caroline Brard

► **To cite this version:**

Caroline Brard. Approche Bayésienne de la survie dans les essais cliniques pour les cancers rares. Santé publique et épidémiologie. Université Paris Saclay (COmUE), 2018. Français. NNT : 2018SACLS474 . tel-02388500

HAL Id: tel-02388500

<https://theses.hal.science/tel-02388500>

Submitted on 2 Dec 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Approche Bayésienne de la survie dans les essais cliniques pour les cancers rares

Thèse de doctorat de l'Université Paris-Saclay
préparée à l'Université Paris-Sud

Ecole doctorale n°570 École Doctorale de Santé Publique (EDSP)
Spécialité de doctorat : Santé Publique - Biostatistiques

Thèse présentée et soutenue à Villejuif, le 20 novembre 2018, par

CAROLINE BRARD

Composition du Jury :

Pr Jean-Yves Blay PU-PH, Université Lyon Bernard	Président
Dr Raphaël Porcher MCU-PH, Université Paris Descartes	Rapporteur
Dr Sarah Zohar DR, INSERM U1138 Équipe 22	Rapporteur
Dr David Dejardin PhD, Laboratoire Roche	Examineur
Pr Chantal Guihenneuc-Jouyaux PU, Université Paris Descartes	Examineur
Pr Laurence Meyer PU-PH, Université Paris Saclay	Examineur
Dr Marie-Cécile Le Deley MCU-PH, Université Paris Saclay	Directeur de thèse
Dr Gwénaél Le Teuff PhD, Université Paris Saclay	Co-encadrant de thèse

Avant-propos

Cette thèse a été préparée au sein du service de biostatistique et d'épidémiologie de Gustave Roussy dirigé par le Dr Ellen Benhamou, et au sein de l'équipe INSERM ONCOSTAT du Centre de recherche en Epidémiologie et Santé des populations (CESP) dirigée par le Dr Stefan Michiels.

Ce travail a été réalisé sous la direction du Dr Marie-Cécile Le Deley et du Dr Gwénaél Le Teuff, chercheurs de l'équipe INSERM ONCOSTAT du Centre de recherche en Epidémiologie et Santé des Populations, Gustave Roussy, en collaboration avec le Dr Lisa Hampson, associate professor à l'Université de Lancaster et chercheur biostatisticienne pour le laboratoire Novartis.

Cette thèse a bénéficié d'un financement du ministère de l'enseignement supérieur et de la recherche, de la ligue contre le cancer et de l'INSERM.

Remerciements

Je remercie le Pr. Jean-Yves Blay d'avoir accepté de présider mon jury de thèse. Je remercie également grandement le Dr. Sarah Zohar et le Dr. Raphaël Porcher d'avoir accepté d'être rapporteurs de ma thèse. Je les remercie pour tous leurs précieux commentaires et suggestions. Enfin, ma gratitude s'adresse également au Dr. David Dejardin, au Pr. Chantal Guihenneuc-Jouyaux et au Pr. Laurence Meyer pour avoir accepté d'examiner ce travail.

Je tiens à remercier le Dr. Marie-Cécile Le Deley et le Dr. Gwénaél Le Teuff d'avoir dirigé ce travail pendant quatre ans. Travailler sous leur direction m'a beaucoup apportée, notamment grâce à la complémentarité de leurs compétences. J'ai beaucoup apprécié leur rigueur et je les remercie de m'avoir fait partager leur expertise scientifique tout au long de cette aventure.

Mes remerciements vont également au Dr. Lisa Hampson pour m'avoir largement épaulé tout au long de ce travail. Je la remercie pour le temps qu'elle m'a consacré, pour ses conseils et son aide précieuse.

Je remercie le Dr. Ellen Benhamou de m'avoir accueilli au sein du service de biostatistiques et d'épidémiologie de Gustave Roussy, ainsi que le Dr. Stefan Michiels au sein de l'équipe Oncostat - INSERM U1018 du CESP.

J'adresse également mes remerciements à l'ensemble du service de biostatistique et d'épidémiologie de Gustave Roussy pour leur accueil chaleureux. J'ai beaucoup apprécié la bonne humeur et la bonne ambiance qui règne au sein du service. Un grand merci à tous les juniors qui y rendent l'ambiance si agréable.

Un très grand merci à mon bureau. C'est en grande partie grâce à votre soutien que j'arrive au bout de ce long chemin! Raïssa, et Caroline, merci pour vos conseils et votre écoute

dans les bons et les mauvais moments. Benjamin et Maritina, vous êtes avec moi dans cette aventure depuis l'origine. Votre aide et votre amitié ont été très précieux pour moi pendant cette thèse. Je vous remercie pour m'avoir soutenu pendant ces quatre ans.

Enfin, je remercie mes parents et mes frères pour leur soutien inébranlable. Merci d'avoir cru en moi jusqu'au bout.

Merci également à toi Rémy d'avoir été là chaque jour, et de m'avoir soutenu (supporté).

Tu as été mon pilier pendant toutes ces années.

Je dédis bien évidemment cette thèse à Arthur, mon plus bel accomplissement.

Liste des productions scientifiques

Articles directement issus du travail de thèse

Brard C, Le Teuff G, Le Deley M-C, Hampson L.V. Bayesian survival analysis in clinical trials : What methods are used in practice? *Clinical Trials*, 2017 Feb ; 14(1) :78-87.
doi : [10.1177/1740774516673362](https://doi.org/10.1177/1740774516673362)

Brard C, Hampson L.V, Gaspar N, Le Deley M-C, Le Teuff G. Incorporating individual historical controls and aggregate treatment effect estimates into a Bayesian survival trial : a simulation study. [Soumis au journal BMC-MRM]

Brard C, Piperno-Neumann S, Delaye J, Brugière L, Hampson L.V, Le Teuff G, Le Deley M-C, Gaspar N. Sarcome-13/OS2016 trial protocol : a Multicentre, randomised, open-label, phase II trial of mifamurtide combined with post-operative chemotherapy for newly diagnosed high risk osteosarcoma patients. [Soumis au journal BMJ open]

Communications orales

Brard C, Le Teuff G, Hampson L.V, Le Deley M-C. Bayesian survival analysis in clinical trials : What methods are used in practice? A systematic review of the literature. *36th Conference of the International Society for Clinical Biostatistics (ISCB)*, Août 2015, Utrecht (Pays-Bas).

Brard C, Hampson L.V, Le Deley M-C, Le Teuff G. Designing a Bayesian randomised controlled trial in osteosarcoma : How to incorporate historical data? *37th Annual meeting of the Society for Clinical Trials (SCT)*, Mai 2016, Montréal (Canada).

Brard C, Hampson L.V, Le Deley M-C, Le Teuff G. Planification d'un essai randomisé Bayésien : spécification du modèle incorporant les données historiques. *EPICLIN 10 / 23^e Journées des statisticiens des CLCC*, Mai 2016, Strasbourg (France).

Brard C, Hampson L.V, Le Deley M-C, Le Teuff G. Planification d'un essai randomisé Bayésien : spécification du modèle incorporant des données historiques individuelles et agrégées. *Conférence AppliBUGS*, Décembre 2016, Paris (France).

Communication affichée

Brard C, Hampson L.V, Le Deley M-C, Le Teuff G. Incorporating individual historical control and aggregated treatment effect into Bayesian survival clinical trial : a simulation study *38th Conference of the International Society for Clinical Biostatistics (ISCB)*, Juillet 2017, Vigo (Espagne).

Abréviations

ASTERIX	Advances in Small Trials dESign for Regulatory Innovation and eXcellence
BMC-MRM	BioMed Central Medical Research Methodology
BMJ	British Medical Journal
EFS	Event-free survival (survie sans événement)
EMA	European Medical Agency
ESMO	European Society for Medical Oncology
FDA	Food and Drug Administration
HR	Hazard Ratio
IC95%	Intervalle de confiance à 95%
ICr95%	Intervalle de crédibilité à 95%
IDeAl	Integrated Design and Analysis of small population group trial
InSPiRe	Innovation in Small Populations Research
ISCB	International Society for Clinical Biostatistics
Log(HR)	Logarithme du hazard ratio
MAMS	Multi-arm Multi-stage (Multi-bras Multi-étape)
MAP prior	Meta-analytic predictive prior
MCMC	Markov Chain Monte Carlo (Monte Carlo par chaînes de Markov)
OS	Overall Survival (survie globale)
PHRC-K	Programme Hospitalier de Recherche Clinique en cancérologie
RCE	Rare Cancers Europe
RMSE	Root Mean Square Error (Racine de l'erreur moyenne quadratique)
SCT	Society for Clinical Trials
SD	Standard Deviation (Écart-type)
SE	Standard Error (Erreur standard)

Table des matières

1	Introduction générale	1
1.1	Évaluation thérapeutique dans les pathologies rares	1
1.1.1	Contexte	1
1.1.2	Approches proposées pour les essais dans les pathologies rares	2
1.2	L'approche Bayésienne dans les essais cliniques comparatifs	10
1.2.1	Principe	10
1.2.2	Intérêts de l'approche Bayésienne	13
1.2.3	Distributions a priori et méthodes Bayésiennes	15
1.2.4	Expression des résultats et règles de décision	20
1.2.5	Difficultés de l'approche Bayésienne	21
1.3	Objectifs de la thèse	23
2	L'analyse Bayésienne de la survie dans les essais cliniques : revue de la littérature	25
2.1	Bases de données et critères de sélection	25
2.1.1	Critères de sélection	25
2.1.2	Organisation de la recherche	26
2.2	Définition des données extraites des publications	28
2.3	Résultats	30
2.3.1	Sélection des articles	30
2.3.2	Caractéristiques générales des essais	32
2.3.3	Design et conduite des essais	34
2.3.4	Modélisation Bayésienne de la survie	36

2.3.5	Choix de la distribution a priori	39
2.3.6	Règles de décision	40
2.3.7	Qualité et exhaustivité des résultats de l'analyse Bayésienne de la survie	41
2.3.8	Maladies rares	43
2.4	Discussion	44
3	Proposition d'un design Bayésien avec incorporation de données historiques individuelles et agrégées pour un essai clinique	49
3.1	Essai Sarcome-13	50
3.1.1	Évaluations pré-existantes du mifamurtide dans l'ostéosarcome	51
3.1.2	Mise en place de l'essai Sarcome-13	52
3.1.3	Données historiques individuelles	53
3.1.4	Données historiques agrégées	55
3.2	Construction des distributions a priori	56
3.2.1	Power prior : Données individuelles sur le bras contrôle	56
3.2.2	Mixture prior : Données agrégées sur l'effet du traitement	58
3.2.3	Distribution a priori jointe	60
3.3	Étude de simulations	61
3.3.1	Définition des données individuelles historiques contrôles	61
3.3.2	Génération des données du nouvel essai	64
3.3.3	Stratégie d'analyse des données simulées	69
3.3.4	Critères d'évaluation de l'impact des données historiques sur l'effet du traitement estimé	71
3.4	Résultats	72
3.4.1	Impact de l'inclusion de données historiques agrégées portant sur l'effet du traitement seulement	72
3.4.2	Impact de l'inclusion de données historiques individuelles portant sur le bras contrôle seulement	81
3.4.3	Incorporation de données historiques individuelles contrôles et de données historiques agrégées sur l'effet du traitement	84

Table des matières

3.4.4	Impact de l'inclusion de données historiques en cas de conflit en matière de distributions de survie	88
3.4.5	Simulations intégrant des données réelles historiques	89
3.5	Discussion	92
4	Expression des résultats d'une analyse Bayésienne de la survie d'un essai clinique randomisé	99
4.1	Introduction	99
4.2	Matériels et méthodes	100
4.2.1	Applications à des essais randomisés réels en oncologie	100
4.2.2	Plan de l'analyse Bayésienne	104
4.3	Résultats	107
4.3.1	Analyse Bayésienne avec distribution a priori non-informative	107
4.3.2	Analyse Bayésienne avec distributions a priori archétypales : analyse de sensibilité	114
4.4	Discussion	119
5	Discussion générale	123
5.1	Synthèse des résultats et apport de la thèse	123
5.2	Limites	126
5.3	Conclusion et perspectives	129
	Bibliographie	133
	Annexe A Chapitre 2: Catégories de Web of Science utilisées	151
	Annexe B Chapitre 2: Spécialités médicales et journaux considérés pour la seconde partie de la recherche	153
	Annexe C Chapitre 2: Formulaire d'extraction de données	157
	Annexe D Chapitre 2: Détails de l'analyse Bayésienne rapportée dans les 28 articles inclus dans la revue de la littérature	165

Annexe E Chapitre 3 : Représentations graphiques des différents scénarios	169
Annexe F Chapitre 3 : Impact de conflits en matière de paramètres entre les données historiques et actuelles	173
Annexe G Chapitre 3 : Impact des données historiques pour différentes variances de la composante informative	181
Annexe H Chapitre 3 : Impact d'un conflit en matière de distributions de survie entre les données historiques contrôle et les données actuelles	189
Annexe I Chapitre 3 : Impact des données historiques contrôles réelles, scénarios S33-S44	195

Liste des Tableaux

2.1	Caractéristiques générales des 28 essais identifiés	33
2.2	Caractéristiques du design et de la conduite des 28 essais identifiés	35
2.3	Spécification de l'analyse de survie Bayésienne, globalement et selon son objectif, pour les 28 essais identifiés	38
2.4	Rapport et caractérisation des distributions a priori et a posteriori	42
3.1	Paramètres des modèles de Weibull et exponentiel par morceaux selon le jeu de données historiques utilisé	62
3.2	Résumé des 32 scénarios considérés pour l'étude de simulations	68
3.3	Modèle exponentiel par morceaux : impact de α_0 en termes de biais pour différentes tailles d'échantillon (scénario S3)	84
3.4	Impact d'un conflit en matière de distribution de survie entre les données historiques et les actuelles pour le scénario S27 en fonction de α_0	89
3.5	Résumé des 12 scénarios considérant les données réelles SARC-OS	91
4.1	Résultats fréquentistes des essais réanalysés	103
4.2	Distributions a priori du Log(HR) pour chaque essai	105
4.3	Probabilités a posteriori d'avoir un certain bénéfice pour les essais EE99-R2Loc et IALT	111
4.4	Essai EE99-R2Loc : Probabilités a posteriori d'avoir un certain bénéfice pour différentes valeurs seuils et selon différentes distributions a priori	115

4.5	Essai IALT : Probabilités a posteriori d’avoir un certain bénéfice pour différentes valeurs seuils et selon différentes distributions a priori	117
4.6	Essai HERBY : Probabilités a posteriori d’avoir un certain effet du traitement pour différentes valeurs seuils et selon différentes distributions a priori	118

Liste des Figures

2.1	Diagramme de flux de la revue systématique de la littérature	31
3.1	Survie sans événement dans le sous-groupe de patients de l'essai historique OS2006	54
3.2	Distribution du <i>mixture prior</i> de l'effet du traitement pour différentes valeurs de ω	60
3.3	Données historiques individuelles sur le bras contrôle considérées dans l'étude de simulations	63
3.4	Survie sans événement des données historiques et des données du bras contrôle du nouvel essai selon leur commensurabilité	66
3.5	Effet de la variance de la composante vague du <i>mixture prior</i> sur la pondération des données historiques sur l'effet du traitement dans le scénario S3	70
3.6	Impact de α_0 et ω sur les caractéristiques opératoires pour les scénarios S1-S4	75
3.7	Impact de α_0 et ω sur les caractéristiques opératoires pour les scénarios S5-S8	76
3.8	Impact de α_0 et ω sur les caractéristiques opératoires pour les scénarios S9-S12	77
3.9	Impact de ω , pour différentes variances de la composante informative, pour les scénarios S1-S4	80
3.10	Distribution a posteriori du Log(HR) selon différentes valeurs des paramètres de pondération	87
4.1	Distribution a posteriori empirique du Log(HR) de l'essai EE99-R2Loc, estimé par un modèle de Cox Bayésien	107

4.2	Représentation de la probabilité a posteriori de l'effet du traitement P(Log(HR) < 0), dans l'essai EE99-R2Loc	108
4.3	Distribution a posteriori empirique du Log(HR) de l'essai IALT, estimé par un modèle de Cox Bayésien	109
4.4	Représentation de la probabilité a posteriori de l'effet du traitement P(Log(HR) < 0), dans l'essai IALT	110
4.5	Distribution a posteriori empirique du Log(HR) de l'essai HERBY, estimée par un modèle de Cox Bayésien	112
4.6	Représentation de la probabilité a posteriori de l'effet du traitement P(Log(HR) < 0), dans l'essai HERBY	113
4.7	Représentation de la distribution a priori, de la vraisemblance, et de la distribution a posteriori de l'effet traitement dans l'essai EE99-R2Loc	114
4.8	Représentation de la distribution a priori, de la vraisemblance, et de la distribution a posteriori de l'effet traitement dans l'essai IALT	116
4.9	Représentation de la distribution a priori, de la vraisemblance, et de la distribution a posteriori de l'effet traitement dans l'essai HERBY	117

— Chapitre 1 —

Introduction générale

1.1 Évaluation thérapeutique dans les pathologies rares

1.1.1 Contexte

Aujourd'hui, on estime que près de 20% des cancers correspondent à des cancers rares (définis comme ayant une prévalence < 5 cas pour 10 000 ou une incidence 6 pour 100 000 par an en Europe) [1–3]. Ces derniers représentent un défi pour les statisticiens en particulier dans l'évaluation de nouveaux traitements en raison du faible nombre de patients disponibles, et par conséquent, de l'incertitude quant aux résultats observés. Cette difficulté est ancienne pour un certain nombre de sous-spécialités oncologiques, telle que la cancérologie pédiatrique par exemple. En effet, chaque type de cancer de l'enfant est un cancer rare. Cependant, cette problématique est de plus en plus d'actualité en oncologie médicale avec le développement de la médecine personnalisée et des thérapies dites "ciblées", qui conduisent à développer des essais d'efficacité sur des groupes de patients de plus en plus restreints.

Sur la base des travaux de Fisher [4], il est habituel dans les essais de Phase III de fixer le risque de faux positifs (erreur de type I, notée alpha) à 5% de formulation bilatérale. Dans le contexte des pathologies rares, malgré les collaborations multicentriques voire

internationales, planifier un essai thérapeutique visant à évaluer et tester l'efficacité d'un nouveau traitement dans un temps raisonnable selon une approche classique fréquentiste, conduit à des compromis en matière de risques d'erreur. En particulier, en conservant un risque de faux positifs au niveau habituel, il peut s'avérer impossible d'assurer une puissance suffisante pour un effet du traitement raisonnable, et pourtant cliniquement pertinent [5]. De plus, l'interprétation des résultats d'un essai est souvent réduite à une conclusion binaire, classiquement basée sur la p-valeur, conduisant trop souvent à la conclusion d'un essai "négatif" sans aucune précaution, quand la p-valeur observée est supérieure au seuil alpha fixé. Ceci est particulièrement préjudiciable dans le cadre des pathologies rares, compte tenu du manque de puissance. Récemment, plusieurs groupes d'experts ont insisté sur la nécessité d'évaluer le bénéfice clinique d'une nouvelle thérapeutique à l'aune de la taille d'effet plus que sur le degré de signification de la différence observée par rapport au traitement standard [6–8].

Avant de développer le principe de l'approche Bayésienne, qui fait l'objet de ce travail de recherche, nous évoquerons différentes approches proposées pour le développement d'essais cliniques dans le cadre des pathologies rares, l'approche Bayésienne pouvant éventuellement être combinée aux autres approches.

1.1.2 Approches proposées pour les essais dans les pathologies rares

De nombreux travaux de recherche portent sur la question de l'évaluation thérapeutique dans les pathologies rares. Ainsi, en 2013, la commission européenne a financé trois projets de recherche internationale afin de développer des méthodologies innovantes et d'établir des recommandations pour la recherche sur des populations de petite taille :

- Le projet IDEAl (*Integrated Design and Analysis of small population group trial*) [9] a pour but d'explorer de nouvelles méthodes pour la conception et l'analyse des essais cliniques, et de formuler une stratégie efficace pour l'évaluation des essais dans les pathologies rares.

- Le projet InSPiRe (*Innovation in Small Populations Research*) [10] est divisé en quatre domaines : les essais de recherche de dose en phase précoce, la théorie de la décision, les essais confirmatoires de petite taille, et la médecine personnalisée.
- Le projet ASTERIX (*Advances in Small Trials dEsign for Regulatory Innovation and eXcellence*) [11] a pour objectif de développer et valider des designs innovants pour des essais cliniques coût-efficaces.

Ces projets ont émis des recommandations sur l'utilisation de certains designs et méthodes, permettant de surmonter les difficultés statistiques liées aux petits échantillons [12]. Les méthodes évoquées regroupent notamment les essais non-randomisés, les essais enrichis, les designs adaptatifs, et l'approche Bayésienne [9–11]. La *Food and Drug Administration* (FDA) a par ailleurs approuvé l'utilisation d'expérimentations *in silico* chez l'animal, et semble favorable à leur utilisation chez l'homme en complément des essais traditionnels [13]. L'initiative européenne *Rare Cancers Europe* (RCE) a également évoqué la possibilité d'utiliser des critères de substitution, ou de relâcher les erreurs de type I et II [14]. Ces différentes approches sont résumées ci-dessous, à la fois pour élargir la présentation du contexte et pour préciser les options que nous avons retenues dans le cadre de mon travail de recherche.

Essais *in silico*

Les essais *in silico* (par opposition aux expérimentations *in vitro* et *in vivo*) sont des modèles mathématiques simulant l'effet d'un traitement sur un groupe de patients pour une pathologie donnée sur la base des connaissances acquises sur les mécanismes d'actions. Il est ainsi possible de construire une réalité virtuelle et de réaliser des expérimentations *in silico*. La génération d'un effet du traitement sur une population virtuelle peut être obtenue après modélisation des mécanismes de progression de la maladie, et de l'effet de la molécule sur les différents processus.

Ces essais peuvent représenter un complément intéressant afin d'accélérer le développement de nouvelles molécules, et de réduire les essais chez l'homme. Ce type

d'approche pourrait s'avérer particulièrement utile pour les essais dans les pathologies rares pour lesquels il est difficile de recruter des patients [15, 16]. Cela permettrait également de faciliter le choix du design le plus approprié en fonction de la pathologie, du critère de jugement, et du traitement étudié.

Cependant, ce type d'essais est limité par le fait que tout ce qu'un modèle mathématique peut prédire provient des connaissances que nous avons utilisées pour le construire. Par conséquent, aucun résultat inattendu ne peut sortir d'une expérience *in silico*, contrairement aux essais cliniques réels qui se confrontent à la réalité d'observations aléatoires.

Dans le cadre de ce travail de recherche, nous ne considérerons que des études cliniques réelles incluant des patients.

Essais non-randomisés

Dans les pathologies rares, lorsque le recrutement d'un grand nombre de patients est impossible, un essai non-randomisé peut être implémenté [5]. Ces essais sont particulièrement exposés aux biais de sélection. Ils peuvent poser des problèmes d'interprétation liés à une comparaison implicite historique qui n'est pas toujours fiable.

Baey et al. [17] montrent comment la mauvaise spécification du taux de réponse attendu sous l'hypothèse nulle (H_0), dans le cadre d'un essai de phase II selon un plan de Simon, conduit à un mauvais contrôle des erreurs de type I et II. Ainsi, une erreur absolue de 5% dans la détermination de ce taux de réponse attendu, conduit à un taux de faux positifs de 30% lorsqu'il est fixé à 10% dans la planification.

De nombreux experts ainsi que le consortium ASTERIX insistent aujourd'hui sur la nécessité de s'orienter, lorsque cela est possible, vers des essais randomisés, même dans les situations de petits échantillons [11, 18].

Nous avons pris l'option, dans le cadre de ce travail de recherche, de ne considérer que des essais comparatifs randomisés, estimant qu'ils apportent la meilleure estimation de l'efficacité de nouveaux traitements.

Designs adaptatifs

Ce type de designs permet de modifier en cours d'essai certains aspects de l'essai afin de l'optimiser sur la base de l'analyse des données déjà accumulées [19]. Le développement de nouveaux traitements pourrait ainsi être facilité et accéléré [14]. Des techniques statistiques particulières sont nécessaires afin de gérer ce type d'adaptation. Le projet IDEAl, tout en reconnaissant les avantages de ces designs dans le cadre des petits échantillons, a mis en garde la communauté scientifique sur le risque d'inflation incontrôlée de l'erreur de type I [9]. De plus, leur utilisation devrait être explicitement énoncée et détaillée dans le protocole de l'étude.

Parmi les nombreux designs adaptatifs possibles, le design "multi-bras multi-étapes" (MAMS, *Multi-arm multi-stage*) a été proposé pour évaluer simultanément, dans le cadre d'un essai de Phase II/III, plusieurs traitements par rapport à un traitement standard. Cela permet ainsi de réduire le nombre de patients nécessaires par rapport à une approche traditionnelle où chaque nouveau traitement est évalué séparément à travers un essai propre. La mise en place d'un design multi-étapes permet d'arrêter un (ou des) bras de traitement à chacune des analyses intermédiaires pour poursuivre l'essai sur les bras les plus prometteurs d'après des règles prédéfinies [20-23].

D'autres types d'essais de phase II randomisé de sélection ont été proposés parmi lesquels le design *Pick the winner* visant à sélectionner quel traitement, parmi une série d'agents, devrait être testé dans le cadre d'un essai de phase III [24]. Le but n'est donc pas de comparer les traitements entre eux, mais de choisir le traitement le plus prometteur. Ce type d'essai peut permettre de réduire le nombre de patients nécessaires pour l'évaluation de traitements. En revanche, plus le nombre de bras augmente, plus la probabilité de sélectionner le meilleur bras diminue. En d'autres termes, la puissance de l'essai diminue. L'erreur de type I est aussi généralement élevée.

Dans le cadre de cette thèse, nous nous sommes limités à l'évaluation d'essais randomisés en deux groupes parallèles, qu'ils soient de Phase II ou de Phase III, en considérant essentiellement l'approche Bayésienne pour l'analyse finale et non pour les analyses intermédiaires éventuelles.

Essais ciblés et essais enrichis

Lors de la planification de l'essai, mais également en cours d'essai dans le cadre d'essais adaptatifs, il est possible de sélectionner un sous-groupe de patients pour lequel la détection d'un effet du traitement est plus probable que dans la population globale (stratégie d'enrichissement) [25, 26].

Ce type de design permet d'augmenter la puissance de l'essai. Le consortium InSPiRe [10], ainsi que l'initiative RCE [14], évoquent notamment l'intérêt de cette approche dans le cadre de la médecine personnalisée.

Il est cependant difficile d'identifier précisément et de façon fiable les "bons répondeurs" au cours d'un essai, et d'appliquer cette sélection au cours de l'essai lui-même, sans validation.

Nous avons considéré dans ce travail de recherche que la population de l'essai était définie a priori, au mieux en fonction des connaissances sur la question posée, sans évoquer un possible enrichissement en cours d'essai.

Critères de substitution

En oncologie, le critère de jugement reconnu par les autorités est la survie du patient (survie globale, OS)[27]. Cependant, ce critère nécessite souvent un suivi prolongé, et un grand nombre de patients afin d'obtenir une puissance suffisante. Si un allongement de la survie reste le critère de choix en termes de signification clinique, d'autres critères de jugement se sont développés. Parmi les nombreux critères de jugement alternatifs à la survie globale [28], plusieurs critères censurés sont utilisés tels que, selon le contexte de l'essai, la survie sans progression, la survie sans maladie ou sans rechute, la survie sans événement, qu'ils soient ou non supposés jouer un rôle de critère de substitution de la survie globale.

Ces critères de substitution ont l'avantage de pouvoir être mesurés plus précocement et/ou plus facilement [5, 14, 29]. Néanmoins, ces critères nécessitent une validation rigoureuse, démontrant notamment leur bonne corrélation avec le critère final, avant leur utilisation en tant que critères de substitution. Cette phase de validation est particulièrement délicate dans le cadre des pathologies rares puisqu'elle nécessite de nombreuses données.

Il est cependant possible de réaliser des essais en considérant des critères intermédiaires en raison de leur pertinence clinique, bien que l'on n'ait pas démontré qu'il s'agissait de critères de substitution.

Les questions abordées dans mon travail de recherche ne porteront pas sur le choix du critère de jugement. En revanche, nous ne considérerons que des critères censurés (survie globale, survie sans progression, survie sans événement). Dans une revue publiée en 2010, les critères censurés restaient prédominants pour obtenir une autorisation de mise sur le marché en oncologie. Ainsi sur 53 autorisations délivrées par la FDA entre Juillet

2005 et décembre 2007, 30 reposaient sur une analyse d'un critère censuré [30]. De plus, les recommandations actuelles de l'*European Society for Medical Oncology* (ESMO) pour évaluer la magnitude bénéfique clinique reposent avant tout sur des critères censurés : survie globale, survie sans maladie, ou survie sans progression en fonction du contexte [7, 31].

Relâchement des erreurs de type I et II

Dans le cadre habituel des essais fréquentistes, la conclusion est basée sur des tests d'hypothèses. L'erreur de type I (notée alpha) est la probabilité de rejeter l'hypothèse nulle (H_0) d'absence d'effet alors que celle-ci est vraie. Dans le cadre des essais de phase III, cette erreur est généralement fixée à 5% de formulation bilatérale suite aux travaux de Fisher [4]. L'erreur de type II (notée bêta) est la probabilité de ne pas rejeter H_0 alors qu'une certaine hypothèse alternative (H_1) est vraie. Il est souvent accepté que cette erreur soit plus importante que l'erreur de type I. Elle est généralement fixée à 10 ou 20%, correspondant à une puissance de 90 ou 80%.

Afin de présenter des risques acceptables, il est fréquent de fixer un risque d'erreur de type II à 10 ou 20% pour une hypothèse alternative trop optimiste dans les essais portant sur des pathologies rares, ce qui revient à accepter de façon implicite une puissance modérée pour une hypothèse alternative plus raisonnable.

Cependant, maintenir le risque alpha à 5% et accepter une erreur de type II plus élevée, ou de façon équivalente cibler un effet du traitement trop optimiste, expose au risque de passer à côté d'un traitement dont l'efficacité est de moindre magnitude mais pourtant cliniquement pertinente [14]. Dans le contexte des pathologies rares, plusieurs auteurs ont prôné un relâchement de l'erreur de type I pour améliorer la puissance d'un essai [5, 32].

Ces recommandations sont soutenues par les travaux de Sposto et al. [33], et Le Deley et al. [34], ouvrant la voie à une nouvelle manière d'évaluer le design des essais. Dans leurs

travaux, l'évaluation des caractéristiques opératoires de l'essai n'est pas centrée sur les erreurs de type I et II, mais sur le bénéfice moyen de survie attendu. Les auteurs concluent que la réalisation consécutive de plusieurs essais randomisés de petite taille, acceptant une erreur de type I relâchée, conduit à un meilleur gain de survie à long terme que des essais traditionnels de grande taille, conçus pour maintenir des critères strictes avec un effet attendu faible. Les principales conclusions des auteurs ont été confirmées récemment par Bayar et al. [35].

Cependant, ils attirent notre attention sur le fait que les recommandations ne peuvent être basées sur des tendances générales, mais devraient plutôt être formées au cas par cas en fonction des spécificités de la pathologie considérée.

La question des règles de décision et donc du risque de conclusion erronée sera largement évoquée dans ce travail.

Approche Bayésienne

Plusieurs auteurs défendent l'intérêt de l'approche Bayésienne dans le contexte des essais sur des pathologies rares [14, 29, 36–40]. Sa capacité à intégrer des connaissances a priori sur l'effet du traitement dans l'analyse de l'essai, et à exprimer les résultats sous la forme de distributions de probabilités, rend cette approche particulièrement intéressante dans ce contexte [5]. Le consortium ASTERIX [11, 41, 42] évoque en particulier la méthode du *power prior*, développée par Ibrahim et al. [43], permettant d'intégrer de manière pondérée des données historiques provenant de précédents essais dans l'analyse du nouvel essai. Récemment, le consortium InSPiRe [10] a proposé une méthode afin d'incorporer des opinions d'experts a priori dans l'analyse de l'essai [44].

Comme la revue de Biswas et al. [45] l'a rapporté, les méthodes Bayésiennes sont déjà bien implémentées dans la pratique dans les essais cliniques de phase I de recherche de dose. Cependant, cela ne semble pas être le cas pour les essais cliniques de phases II et III, où l'on cherche à évaluer l'efficacité d'un traitement.

1.2 L'approche Bayésienne dans les essais cliniques comparatifs

1.2.1 Principe

L'approche Bayésienne permet de prendre en compte une information externe aux données de l'essai, exprimée sous la forme d'une distribution de probabilité a priori (ou *prior*) de l'effet du traitement $P(\beta)$, que l'on combine aux données de l'essai à l'aide de la fonction de vraisemblance $P(X|\beta)$, et que l'on divise par la probabilité marginale $P(X)$, afin d'estimer une distribution de probabilité a posteriori de l'effet du traitement $P(\beta|X)$. En d'autres termes, le principe de l'analyse Bayésienne permet d'actualiser l'information disponible sur l'étendue plausible des valeurs de l'effet traitement avant l'essai (exprimée par une probabilité dite a priori, ou *prior*) par les données recueillies au cours de l'essai (caractérisées par la fonction de vraisemblance) pour produire une estimation révisée de l'étendue plausible des valeurs de l'effet traitement (exprimée par une probabilité dite a posteriori). La largeur (ou variance) de la distribution a priori représente le niveau de certitude de l'effet traitement. Augmenter la variance de cette distribution correspond à étendre les valeurs possibles de l'effet traitement. L'aire sous la courbe à gauche d'une valeur donnée (distribution cumulée) représente la probabilité que l'effet traitement soit inférieure à cette valeur.

Cela se traduit par la formule ci-dessous issue du théorème de Bayes [46] :

$$P(\beta|X) = \frac{P(X|\beta)P(\beta)}{P(X)}$$

Quand l'effet du traitement est évalué sur un critère censuré (données de survie), ce qui est le cas dans ce travail, le paramètre d'intérêt β est défini comme l'effet relatif du traitement, estimé par le logarithme du rapport des risques instantanés (ou logarithme du *hazard ratio*, noté $(\text{Log}(\text{HR}))$).

Le calcul de la distribution a posteriori peut se faire de manière analytique lorsque les lois a priori et a posteriori sont de la même famille. On parle alors de loi a priori conjuguée, comme nous l'exposerons pour l'analyse de la survie en supposant une loi normale pour le Log(HR) (Chapitre 4, Section 4.2.2) [47]. Cette approche "hybride" consiste à mettre à jour la distribution a priori du Log(HR), qui peut être extrait d'une publication, avec l'estimation du Log(HR) obtenue à partir de l'analyse fréquentiste des données de l'essai. Cette approche présente l'intérêt d'être simple et de ne pas nécessiter de données historiques individuelles. Cependant, elle se limite à l'estimation de la distribution a posteriori de l'effet du traitement intégrant de l'information a priori uniquement sur le Log(HR), et non sur des covariables ou sur la fonction de survie sous-jacente.

Cependant, l'analyse Bayésienne nécessite le calcul d'intégrales. Ce calcul ne peut être réalisé de façon analytique lorsque la loi a priori n'est pas conjuguée. Cela nécessite alors de réaliser soit une approximation par intégration numérique, soit une estimation par simulations de Monte Carlo.

L'intégration numérique fait référence à des méthodes dans lesquelles l'intégrale d'une fonction continue est évaluée en calculant la valeur de la fonction sur un nombre fini de points. En augmentant le nombre de points où la fonction est évaluée, il est possible d'obtenir la précision souhaitée.

Les méthodes stochastiques Monte Carlo par chaînes de Markov (*Markov Chain Monte Carlo*, MCMC) [48] sont des algorithmes d'échantillonnage reposant sur la théorie des chaînes de Markov. Elles sont utilisées pour générer des échantillons tirés d'une loi notée π . Cette loi peut représenter la loi a posteriori d'un ou plusieurs paramètres.

Une chaîne de Markov est une séquence d'échantillon $(x_0, x_1, x_2, \dots, x_n)$ (n : nombre d'échantillons) avec la propriété suivante : la dépendance au passé se limite à la dernière itération. On génère donc un échantillon d'un paramètre, uniquement à partir de l'échantillon précédent, à l'aide d'un opérateur, appelé noyau de la chaîne de Markov,

noté K . Ce noyau est une distribution de probabilité pour passer d'un état à l'état suivant. À partir d'une distribution initiale, on tire une valeur x_0 , et les valeurs suivantes sont tirés de $K(x_i|x_{i-1})$. L'échantillon final ainsi obtenu tend vers la distribution cible π , asymptotiquement ($n \rightarrow \infty$).

Dans ce contexte, les deux algorithmes stochastiques les plus connus sont l'échantillonneur de Gibbs [49, 50] et l'algorithme de Metropolis-Hastings [51, 52].

L'échantillonneur de Gibbs génère une chaîne de Markov en divisant le vecteur de variable θ en sous-vecteurs, et en échantillonnant chaque sous-vecteur conditionnellement à la "valeur la plus récente" des autres éléments de θ . Ainsi, si l'on considère k sous-vecteurs de θ ($\theta = (\theta_1, \theta_2, \dots, \theta_k)$) :

- On choisit arbitrairement des valeurs initiales $\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_k^{(0)}$ restant cohérentes avec la loi a priori, où l'exposant représente le numéro de l'itération (la valeur zéro représentant l'état initial de la chaîne de Markov).
- Pour chaque élément de θ , on échantillonne de nouvelles valeurs de la manière suivante :

On échantillonne une nouvelle valeur pour θ_1 , à partir de la distribution conditionnelle complète de θ_1 en fonction de la valeur la plus récente de tous les autres éléments de θ et des données (notées y) : $\theta_1^{(1)} \sim p(\theta_1|\theta_2^{(0)}, \theta_3^{(0)}, \dots, \theta_k^{(0)}, y)$

On échantillonne une nouvelle valeur $\theta_2^{(1)}$ pour la seconde composante de θ , à partir de sa distribution conditionnelle complète $p(\theta_2|\theta_1^{(1)}, \theta_3^{(0)}, \dots, \theta_k^{(0)}, y)$. Une nouvelle valeur pour θ_1 a déjà été échantillonnée, c'est donc sa valeur "la plus récente" qui est utilisée.

Cette étape se répète pour chacun des sous-vecteurs. Cela complète une itération de l'échantillonnage de Gibbs.

L'algorithme de Metropolis-Hastings [51, 52] échantillonne d'abord des valeurs à partir d'une distribution "proposée", notée $q(\tilde{\theta}|\theta)$. Une probabilité d'acceptation est ensuite calculée pour la valeur proposée ($\tilde{\theta}$). Cette probabilité est convertie en une décision d'acceptation ou de rejet. Si la valeur proposée est acceptée alors la chaîne de Markov passe à cette valeur, sinon elle reste à sa valeur actuelle. La chaîne de Markov peut donc

rester à la même place pour un certain nombre d'itérations, en particulier lorsque la distribution proposée n'approxime pas bien la distribution cible. Malgré le fait que la chaîne contienne plusieurs fois la même valeur, elle peut être utilisée comme si toutes les valeurs avaient été échantillonnées directement, à condition que la chaîne dispose de suffisamment de temps pour explorer pleinement la distribution cible.

Ces méthodes permettent donc de générer des échantillons, dépendants ou non, et identiquement distribués qui convergent plus ou moins rapidement vers la distribution cible π . Il est donc important d'évaluer la convergence des échantillons vers cette distribution et l'autocorrélation entre ces échantillons. Des méthodes graphiques (*Trace plot*, *Autocorrelation plot*, *Running mean plot*, *Q-Q plot*...) et des tests formels (diagnostic de Geweke [53], diagnostic de Heidelberger-Welch [54], diagnostic de Raftery-Lewis [55], diagnostic de Brooks-Gelman-Rubin [56, 57]) existent. En pratique, ces algorithmes d'échantillonnage génèrent habituellement des échantillons avec une première phase dite de "*burn-in*". Afin de minimiser l'effet des valeurs initiales sur l'estimation de la distribution cible cette partie initiale de la chaîne, non stationnaire, est généralement écartée.

Ces méthodes permettent de proposer un modèle de Cox Bayésien ou un modèle paramétrique Bayésien pouvant intégrer de l'information a priori, non seulement sur l'effet relatif du traitement, mais également sur l'effet de covariables et sur la fonction de risque de base en cas de modélisation paramétrique.

1.2.2 Intérêts de l'approche Bayésienne

Nous nous placerons dans le contexte d'un essai clinique avec une perspective opérationnelle : nous souhaitons prendre une décision quant à l'efficacité d'un nouveau traitement en considérant toute l'information disponible.

Dans cette perspective, l'un des aspects les plus intéressants de l'approche Bayésienne est la possibilité de prendre en compte des informations extérieures pour l'analyse du critère de jugement principal de l'essai, au travers de la distribution a priori [58, 59]. Comme détaillé dans la Section suivante 1.2.3, ces informations peuvent être de différents types :

- Opinions d'experts
- Données historiques

L'utilisation d'une distribution a priori non-informative est également possible.

De plus, l'approche Bayésienne permet d'exprimer les résultats sous la forme de probabilités a posteriori et d'intervalle de crédibilité, plus facilement interprétables par le clinicien et le patient qu'une p-valeur ou un intervalle de confiance comme mesures d'incertitude de l'effet du traitement considéré.

Ainsi, le calcul et l'interprétation de la p-valeur correspondent à un raisonnement par l'absurde [60]. Quand on conclut à une différence significative entre deux groupes de traitement au seuil 5%, cela signifie que les données observées sont très "improbables" sous l'hypothèse nulle H_0 et que l'on peut donc rejeter H_0 et conclure à un effet du traitement, au risque de se tromper 5 fois sur 100. Ce concept est loin d'être compris dans la communauté médicale.

L'intervalle de confiance à 95%, qui permet de donner une mesure de la précision de l'estimation, est également souvent interprété à tort comme l'intervalle de valeurs ayant 95 chances sur 100 de contenir la vraie valeur, ce qui correspond stricto sensu à l'intervalle de crédibilité. Expliquer la signification de l'intervalle de confiance est complexe : il correspond à un intervalle tel que, si le même essai était répété 100 fois, 95 des intervalles de confiance obtenus contiendraient la vraie valeur de l'estimation. La simplicité d'interprétation de l'intervalle de crédibilité contraste avec la complexité de l'intervalle de confiance.

Par ailleurs, quand les experts proposent de prendre en compte la magnitude du bénéfice clinique dans la décision, il nous semble que les règles proposées se définissent plus facilement dans le cadre Bayésien (cf. Section 1.2.4).

L'approche Bayésienne peut également être particulièrement intéressante pour "monitorer" les essais. En effet, elle permet d'arrêter l'essai lorsqu'il y a une forte probabilité d'un effet du traitement cliniquement significatif [61, 62], ou au contraire lorsque la probabilité que l'effet du traitement soit cliniquement significatif est faible.

Ces dernières années, l'utilisation des méthodes Bayésiennes a été facilitée par l'amélioration et l'implémentation des techniques de calculs, et la disponibilité d'ordinateurs performants. En 1991, un programme informatique appelé BUGS [63], combinant la modélisation graphique à l'utilisation de l'échantillonnage de Gibbs pour effectuer les calculs, a été développé. Un groupe de programmes appelé WinBUGS est aujourd'hui disponible en ligne. Le logiciel JAGS permet de réaliser l'analyse Bayésienne d'un modèle hiérarchique. De plus, le logiciel SAS dispose à présent de plusieurs macros et procédures, et le logiciel R de nombreux packages (LearnBayes, MCMCpack, bayesSurv...) permettant de réaliser des analyses Bayésiennes. Le langage Stan constitue également un ensemble d'outils pour l'inférence Bayésienne utilisable à travers les logiciels R ou Python par exemple.

1.2.3 Distributions a priori et méthodes Bayésiennes

La construction de la distribution a priori est une étape importante de l'inférence Bayésienne. Il faut en effet caractériser ce qui est connu, ou ce qui est pensé sur l'effet du traitement. Cette distribution peut être ensuite convertie en "pseudo-observations" afin de mieux visualiser ce que représente la quantité d'information contenue dans cette distribution a priori.

La distribution a priori peut être informative ou non-informative. Il existe différents types de distribution a priori informatives et différentes méthodes pour les incorporer.

Distribution a priori informative

La distribution a priori informative peut être constituée d'opinions particulières, sceptiques ou enthousiastes, d'opinions d'un panel d'experts, ou construit à partir de différentes sources telles que des données historiques issues de précédents essais ou études.

Ainsi, des distributions informatives de type archétypales (*from the shelf* en anglais) peuvent être utilisées, choisies pour représenter des opinions sceptiques ou enthousiastes quant à l'effet du traitement [64–67]. Concernant la distribution a priori du Log(HR) dans un essai contrôlé randomisé d'efficacité, la pratique traditionnelle est de définir une distribution a priori sceptique avec une moyenne centrée sur l'absence d'effet du traitement, et une distribution a priori enthousiaste centré sur un HR caractérisant une hypothèse alternative donnée, correspondant le plus souvent à l'hypothèse alternative du calcul du nombre de sujets nécessaire de l'approche fréquentiste. Ainsi, dans le cadre du monitoring d'un essai, on abandonnera un traitement dont l'efficacité s'avèrerait insuffisante, même aux yeux des cliniciens enthousiastes, tandis que l'on conclura à son intérêt si on arrive à convaincre les cliniciens sceptiques [64].

Une distribution a priori peut être également basée sur des opinions d'un panel d'experts interrogés dans le contexte de la problématique considérée, opinions qui sont ensuite transformées en distribution mathématique.

Cette approche nécessite à la fois une bonne pédagogie de la part du statisticien qui recueille les opinions, et une certaine compréhension des statistiques de la part des experts interrogés. Différents processus pour l'élicitation d'opinions d'experts ont été développés [44, 68–77]. Une seule distribution a priori consensuelle peut en être dérivée au

travers d'une agrégation mathématique, ou comportementale (où les experts sont amenés à un consensus ; par exemple la méthode Delphi). Alternativement, un panel d'opinions peut être enregistré pour capturer la diversité des opinions soutenues par la communauté clinique, conduisant à l'élaboration de plusieurs distributions a priori différentes.

Le troisième type de distribution a priori informative repose sur des données dites "historiques", extérieures à l'essai, généralement connues au moment de la planification de l'essai. En effet aujourd'hui, de plus en plus de données sont accessibles aux chercheurs. Elles peuvent être sous forme de données individuelles ou d'informations agrégées, extraites de publications. Elles peuvent provenir d'essais cliniques antérieurs ou être générées par des études observationnelles (données de registres, bases de données médico-administratives...). Elles peuvent porter sur l'effet relatif du traitement expérimental en comparaison au traitement contrôle, ou sur les paramètres d'un des groupes de traitement seulement, en particulier le groupe contrôle. Les informations extérieures sur l'effet possible du nouveau traitement sont généralement utilisées pour justifier l'essai et le calcul du nombre de sujets nécessaire : données d'essai de phase II préalable, données d'un essai réalisé sur une population proche, données d'un essai évaluant un traitement présentant de grandes similitudes avec le traitement évalué par exemple. Il est également fréquent de disposer d'informations sur le traitement servant de traitement de référence pour le groupe contrôle.

Dès 1976, Pocock a suggéré l'intégration de données historiques sur le groupe contrôle dans l'analyse d'un essai randomisé. Il a proposé six critères pour s'assurer de la pertinence des données contrôles historiques par rapport au nouvel essai [78] sans pour autant garantir une parfaite compatibilité des données.

1. Le groupe contrôle historique doit avoir reçu le même traitement standard que le groupe contrôle du nouvel essai.
2. Le groupe contrôle historique doit faire partie d'un essai clinique récent contenant les mêmes critères d'éligibilité pour les patients.
3. La méthode d'évaluation du traitement doit être la même.

4. Les distributions des caractéristiques des patients dans le groupe contrôle historique doivent être comparables avec celles du nouvel essai.
5. L'essai historique doit avoir été réalisé dans la même organisation avec globalement les mêmes investigateurs.
6. Il ne doit pas y avoir d'autres indications laissant penser que les résultats de l'essai historique et du nouvel essai puissent être différents (par exemple un recrutement plus rapide dans la nouvelle étude).

Dans le même esprit, Tan et al. [37] proposent une stratégie pour sélectionner les études à inclure dans la construction de la distribution a priori en tenant compte de la pertinence, de la validité et de la précision. Ces études sont alors pondérées en fonction de ces trois critères.

Qu'il s'agisse de données historiques sur l'effet relatif du traitement évalué ou de données sur l'un des groupes de traitement, la commensurabilité des données historiques avec les données de l'essai actuel est un point de discussion important au moment de la planification et de la réalisation de l'analyse Bayésienne. À noter que le terme de commensurabilité est largement utilisé dans les articles rapportant des méthodes Bayésiennes sans qu'existe une définition formelle. On entend ici par commensurabilité entre deux jeux de données le fait que l'on puisse considérer que ces jeux de données historique et actuel sont générés par des processus statistiques reposant sur des paramètres similaires. Le degré de commensurabilité peut être défini a priori par les experts. Il peut également être évalué au travers de tests.

Dans le contexte des pathologies rares, pour ce qui est de données historiques pouvant enrichir l'information du groupe contrôle, le traitement standard reste souvent stable au cours du temps puisque les avancées vers de nouvelles options sont plus lentes; nous pouvons donc attendre une certaine commensurabilité entre la performance du traitement contrôle dans les données historiques et celle dans le nouvel essai si les critères de sélection des patients et d'évaluation restent stables.

En 2006, la *Food and Drug Administration* (FDA) publie ses recommandations pour l'utilisation des statistiques Bayésiennes dans le cadre des essais cliniques portant sur

les dispositifs médicaux [79]. Ces recommandations soulignent l'avantage apporté par l'intégration de données historiques dans la distribution a priori du paramètre d'intérêt, tout en insistant sur l'importance de pondérer cette information.

Une revue récente a identifié différentes méthodes Bayésiennes et fréquentistes permettant l'incorporation de données historiques dans l'essai actuel [80]. Concernant les méthodes Bayésiennes, les auteurs distinguent les méthodes dynamiques et les méthodes non dynamiques. Dans les méthodes dynamiques, les données extérieures sont pondérées selon leur commensurabilité avec les données actuelles en utilisant un paramètre aléatoire. Dans les méthodes non-dynamiques, le poids accordé aux données extérieures à l'essai est soit pré-spécifié dans le protocole, sur la base d'opinions d'experts ou d'une étude de calibration par simulations, soit fixé au moment de l'analyse au vu de la commensurabilité observée [81].

Parmi les méthodes utilisables pour des critères de survie avec le Log(HR) comme paramètre d'intérêt, une possibilité est d'utiliser un *power prior*, développé par Ibrahim et al. [43] notamment pour les modèles de régression, qui élève à la puissance α_0 la vraisemblance de l'essai historique, afin de pondérer ces données. La valeur de α_0 peut être fixe, ou une distribution a priori peut être définie pour ce paramètre. Dans ce dernier cas, Neuenschwander et al. [82] apportent une correction, proposant ainsi le *modified power prior*, qui comprend une constante d'intégration manquante dans la version originale du *power prior*. Par la suite, Hobbs et al. [83, 84] proposent un *commensurate power prior* pour les modèles hiérarchiques et les modèles linéaires généralisés, introduisant un paramètre supplémentaire, τ , mesurant la commensurabilité entre les données historiques et actuelles. Cette méthode a ensuite été étendue aux données de survie par Murray et al. [85]. Il est en effet possible de construire un modèle hiérarchique Bayésien permettant de prendre en compte la variabilité inter-essais [86]. Cependant, lorsque la distribution a priori est basée sur un ou deux essais seulement, l'estimation de la variabilité inter-essai devient difficile et mène à une estimation de la variance très instable [87]. Une autre approche possible est de formuler un *meta-analytic predictive (MAP) prior* dans le cadre des modèles hiérarchiques afin de combiner des données à travers plusieurs sources de

population hétérogène [88, 89]. Cette méthode peut être approximée par une approche combinant des distributions mixtes conjuguées, appelée *robust MAP prior* [90]. En 2016, Mutsvari et al. [91] étendent cette méthode, qu'ils appellent *mixture prior*, au contexte d'une loi normale.

Distribution a priori non-informative

De nombreux cliniciens et statisticiens, ainsi que les autorités de santé, trouvent difficile d'accepter l'inclusion d'une distribution a priori basée sur des éléments subjectifs [92, 93].

L'utilisation d'une distribution a priori non-informative, où toutes les tailles d'effet du traitement sont équiprobables, peut être proposée pour l'analyse principale ou secondaire d'un essai clinique [5]. La distribution a posteriori coïncide alors avec la vraisemblance, ne reflétant que ce que les données de l'essai apportent à la connaissance. Elle peut également compléter la distribution a priori informative dans le cadre d'une analyse de sensibilité.

1.2.4 Expression des résultats et règles de décision

L'approche Bayésienne permet d'enrichir l'approche fréquentiste classique en fournissant une description complète de la distribution de l'effet d'un traitement, permettant de traduire les résultats dans une forme plus facilement interprétable par les cliniciens et les patients (probabilité d'une certaine quantité d'effet du traitement, ou intervalle de crédibilité), et de choisir d'autres règles de décision que le test habituel de rejet de l'hypothèse nulle (absence d'effet traitement) [5, 94]. À la fin d'un essai, l'incertitude sur les bénéfices et risques d'un traitement peut donc être quantifiée, et ainsi contribuer aux futures prises de décision [95].

Récemment, de nombreux experts ont insisté sur la nécessité d'évaluer l'effet d'un traitement non seulement à l'aune de la significativité du test, mais également de la magnitude de l'effet clinique. Le seuil pour considérer la pertinence clinique peut prendre

en compte le risque de base de la population à l'étude, la complexité de l'intervention évaluée, le risque d'effets indésirables, et les autres critères cliniques [7, 31, 96]. En 2015, puis en 2017, un groupe d'experts de l'ESMO a formalisé ce concept de pertinence clinique [7, 31]. Ces recommandations ont été établies dans un cadre fréquentiste en comparant, à une certaine valeur seuil, la borne inférieure de l'intervalle de confiance à 95% pour prendre en compte ce qu'ils appellent la "variabilité" du HR. Ainsi, pour un essai utilisant la survie globale comme critère de jugement avec une médiane de survie dans le groupe contrôle de plus de 12 mois, les experts concluent à un bénéfice cliniquement pertinent si la borne inférieure de l'intervalle de confiance à 95% est inférieure à 0.7. L'interprétation de cette règle ne nous semble pas immédiate contrairement à une règle Bayésienne basée sur l'expression probabiliste de l'effet du traitement par rapport à un seuil. On pourrait ainsi proposer de considérer comme bénéfique cliniquement pertinent un traitement pour lequel la probabilité d'un $HR < 0.7$ serait supérieure à 2.5% (si l'on reste sur une traduction grossière de la règle fréquentiste évoquée ci-dessus, tout autre seuil pouvant être considéré). D'autres règles plus intuitives pourraient également être proposées.

Dans le cadre du monitoring des essais, des règles d'arrêt Bayésiennes peuvent être proposées pour les analyses intermédiaires basées sur les probabilités a posteriori de l'effet du traitement [64, 97–103].

1.2.5 Difficultés de l'approche Bayésienne

Des difficultés techniques existent dans la construction de la distribution a priori, en particulier quand le critère étudié est une donnée censurée.

Dans l'analyse habituelle fréquentiste des essais cliniques sur données censurées, un modèle de Cox semi-paramétrique est souvent utilisé pour évaluer l'effet relatif d'un traitement par le *hazard ratio* (HR). Il ne nécessite pas de modéliser la fonction de survie sous-jacente (risque de base).

Dans le cadre Bayésien, comme expliqué en Section 1.2.1, il est possible de réaliser une analyse Bayésienne simple que nous avons appelée "hybride". En effet, en considérant que le $\text{Log}(\text{HR})$ suit une loi normale [47], l'utilisation d'une loi a priori conjuguée permet d'intégrer des données historiques sur l'effet relatif du traitement. Cette approche hybride consiste à mettre à jour la distribution a priori, avec l'estimation du $\text{Log}(\text{HR})$ obtenue à partir d'une analyse fréquentiste. La distribution a posteriori du paramètre peut alors être déterminée de manière analytique [69, 104]. Le détail de cette méthode sera exposé Chapitre 4 Section 4.2.2.

En revanche, si l'on souhaite intégrer des informations sur une fonction de survie, par exemple parce que l'on dispose de données historiques individuelles pouvant enrichir l'information sur le groupe contrôle, il devient nécessaire de préciser un processus stochastique particulier, donc de modéliser la fonction de survie; la spécification du modèle le plus adéquat pour le risque de base peut être difficile en particulier dans le contexte des pathologies rares.

Dans le cadre d'essais cliniques visant à évaluer de nouveaux traitements, outre les difficultés techniques pour les statisticiens, la principale difficulté pour les destinataires des analyses (communauté médicale et scientifique, décideurs...) concerne la subjectivité dans l'interprétation de l'essai. Celle-ci est liée à l'intégration d'une distribution a priori informative pour des paramètres inconnus comme l'effet du traitement, les paramètres du groupe contrôle, ou les effets des covariables. Bien qu'elle reconnaisse l'intérêt de l'approche Bayésienne pour les essais cliniques dans les maladies rares, l'*European Medical Agency* (EMA) exprime certaines réserves quant à l'incorporation de croyances a priori, et au fort impact qu'elles peuvent avoir sur les résultats et sur la décision finale corrélée [29]. Même si les distributions a priori sont basées sur des données réelles, l'investigateur doit décider quels jeux de données vont être incorporés et comment les pondérer en fonction de leur commensurabilité anticipée ou observée avec les futures données de l'essai [105–107]. L'article de l'EMA sur l'extrapolation souligne que l'on devrait quantifier l'incertitude sur les différences possibles entre la population source et la

population cible [106]. Lorsqu'il n'existe aucun jeu de données pertinent, les distributions a priori peuvent être basées sur des opinions d'experts. Cependant, il peut être difficile de savoir qui considérer comme un "expert", et la subjectivité de ce type de distribution a priori est fortement critiquée [93, 108].

1.3 Objectifs de la thèse

Le premier axe de travail a consisté en une revue systématique de la littérature afin de faire un état des lieux de l'utilisation des méthodes Bayésiennes dans l'analyse des essais cliniques ayant pour but d'évaluer l'effet d'un traitement sur un critère censuré (Chapitre 2). Cela a permis de mieux appréhender à la fois l'apport de cette approche, les aspects techniques sous-jacents, et également les difficultés soulevées dans des essais réels. Cette revue de la littérature a fait l'objet d'une présentation orale à l'*International Society for Clinical Biostatistics* (ISCB) à Utrecht (Pays-Bas) en août 2015 et a abouti à la publication d'un article scientifique dans le journal *Clinical Trials* [109] en février 2017.

Le second axe de la thèse a porté sur la proposition d'un design Bayésien avec critère de survie, intégrant des données historiques dans le cadre des ostéosarcomes à haut risque de rechute. Nous avons proposé une approche pour intégrer et pondérer à la fois des données historiques individuelles sur le bras contrôle d'un essai, à travers un *power prior* [43], et des données historiques agrégées portant sur l'effet du traitement, à travers un *mixture prior* [91]. Afin d'évaluer les caractéristiques opératoires de cette approche, et d'aborder la problématique de la commensurabilité entre les données historiques et les données du nouvel essai, et l'impact du choix de la modélisation de la survie de base, une étude de simulations a été réalisée considérant différents scénarios. En effet, les problématiques de la commensurabilité et du choix du meilleur modèle pour le risque de base sont exacerbées dans le cadre des pathologies rares (Chapitre 3). Ce travail a été présenté à la *Society for Clinical Trials* (SCT) à Montréal (Canada), au congrès Epiclin à Strasbourg en mai 2016, au

séminaire AppliBUGS à Paris en décembre 2016, et à l'ISCB en juillet 2017. Un article est en cours de révision au journal *BioMed Central Medical Research Methodology* (BMC-MRM). Par ailleurs, le design Bayésien proposé a été intégré au protocole d'un essai, qui est ouvert aux inclusions depuis septembre 2018, et qui a fait l'objet d'un financement dans le cadre du Programme Hospitalier de Recherche Clinique en cancérologie 2016 (PHRC-K 2016). Ce protocole a été soumis pour publication au *British Medical Journal* (BMJ) open.

Enfin, le troisième axe de la thèse a porté sur l'apport de l'approche Bayésienne dans l'expression des résultats, leur interprétation, et la définition des règles de décision, illustré à travers la ré-analyse finale de trois essais randomisés avec critère censuré par rapport à l'analyse initiale fréquentiste (Chapitre 4).

— Chapitre 2 —

L'analyse Bayésienne de la survie dans les essais cliniques : revue de la littérature

Bien que les travaux méthodologiques portant sur l'analyse Bayésienne d'un critère d'efficacité dans les essais cliniques se multiplient [83–85, 88–91, 110], et que les outils de calculs se développent, l'implémentation de l'approche Bayésienne dans la pratique reste mal connue.

Une revue systématique des essais cliniques utilisant une analyse Bayésienne de la survie nous a permis de décrire quelles distributions a priori et méthodes sont utilisées, et comment ces méthodes sont implémentées et leurs résultats rapportés.

2.1 Bases de données et critères de sélection

2.1.1 Critères de sélection

Cette recherche a été réalisée en septembre 2015. Tous les articles publiés en anglais, rapportant des essais cliniques incorporant une analyse Bayésienne de la survie, ont été considérés comme éligibles sans restriction en matière de date de publication.

Afin de sélectionner les articles pertinents pour la revue, des critères de recherche ont été définis. Les articles contenant les termes suivants dans leur titre, résumé ou mots clés ont été considérés éligibles :

(Bayes* OR 'prior distribution*' OR 'posterior distribution*' OR 'credib* interval*')

AND (trial* OR controlled OR randomi?ed OR 'phase*study')

AND (survival OR 'time to' OR 'failure time').

Les études observationnelles, coût-efficacité, et de recherche de doses ont été exclues en ajoutant les conditions de recherche suivantes :

NOT (observational OR epidemio* OR 'case control')

NOT (cost effect*)

NOT (dose finding).

Les ré-analyses d'essais à des fins cliniques ont été considérées éligibles. Cependant, celles réalisées dans le seul but d'illustrer une nouvelle méthode statistique ont été exclues, tout comme les articles de méthodologie statistique et les méta-analyses. Les articles incluant un critère de survie considéré comme un critère binaire sans données censurées ont également été exclus.

2.1.2 Organisation de la recherche

La recherche d'articles a été réalisée en deux parties.

La première partie de la recherche est une recherche automatique qui s'est concentrée sur les bases de données *PubMed* et *Web of Science*. Pour cette dernière, nous avons restreint notre attention aux domaines médicaux et de recherche en statistiques en utilisant le filtre par catégories disponible sur *Web of Science* (Voir Annexe A). Les titres et résumés (et le texte intégral si nécessaire) de tous les articles identifiés par ce processus ont été examinés afin de confirmer leur éligibilité pour la revue systématique.

La seconde partie de cette recherche est une recherche manuelle qui consistait à identifier les articles directement sur les sites internet des journaux pour lesquels une recherche dans le texte intégral était possible. Cela nous a permis de sélectionner les articles évoquant une analyse de survie Bayésienne dans le corps du texte mais pas dans le titre, le résumé, ou les mots clés, et donc ignorés par la recherche automatique. Nous avons sélectionné cinq revues médicales généralistes permettant une recherche dans le texte de leurs articles sur leur site internet et ayant les plus forts facteurs d'impact en 2014. De façon similaire, pour chaque domaine médical, nous avons sélectionné le journal spécialisé ayant le plus haut facteur d'impact et permettant une recherche dans le texte intégral des articles (Annexe B). De plus, les journaux contribuant à au moins un article identifié lors de la première partie de la stratégie de recherche ont été considérés. Au total, nous avons sélectionné 28 revues médicales pour cette recherche (dont cinq en oncologie). Nous avons examiné le texte intégral de tous les articles identifiés par la seconde stratégie de recherche pour confirmer leur inclusion dans la revue systématique.

Un second lecteur a examiné indépendamment une sélection aléatoire de 10% des articles identifiés par les deux parties de la recherche pour vérifier leur éligibilité.

Enfin, cette recherche à partir des données publiées a été complétée par une recherche sur le site *clinicaltrial.gov* afin de chercher les essais en cours ou finalisés, mentionnant une analyse Bayésienne de la survie dans le résumé publié en ligne.

2.2 Définition des données extraites des publications

Un formulaire comprenant 100 items (Annexe C) a été développé afin d'extraire les données suivantes :

- Caractéristiques générales de l'essai (âge des patients recrutés, rareté de la pathologie considérée¹, durée de recrutement des patients, type de traitement évalué)
- Design (phase de l'essai, randomisation, aveugle, design adaptatif ou non, taille de l'échantillon prévue, critères principal et secondaire de l'essai)
- Analyse de survie Bayésienne (objectif de l'analyse Bayésienne, type de critère de jugement utilisé pour l'analyse, règle de décision définie, modèle utilisé, logiciel utilisé pour l'analyse)
- Distributions a priori définies (avis d'experts, distributions a priori archétypales, données historiques, méthodes d'incorporation de l'information a priori, pondération de l'information ou non, analyse de sensibilité réalisée ou non)
- Qualité et exhaustivité des résultats d'analyse Bayésienne (rapport et/ou représentation des distributions a priori et a posteriori)

Les analyses de survie ont été classées selon l'une des approches suivantes :

1. Analyse de survie limitée à un temps donné.
2. Analyse de survie classique utilisant l'ensemble des temps de participation, (appelée ensuite critère de survie "continu"), l'approche Bayésienne pouvant être de deux types :
 - (a) Approche basée sur l'estimation du logarithme du *hazard ratio* (Log(HR)), supposée suivre une distribution normale [47]. Cette approche hybride consiste à mettre à jour la distribution a priori du Log(HR) (loi conjuguée), avec

1. Nous avons évalué la rareté de la pathologie considérée sur la base de la prévalence de la maladie, extraite du site web Orphanet (<http://www.orpha.net/consor/cgi-bin/index.php>). Les maladies ont été classées comme rares si la prévalence estimée était inférieure à 5 pour 10 000 personnes en Europe (définition de l'EMA) [1]

2.2. Définition des données extraites des publications

l'estimation de la vraisemblance partielle de ce paramètre obtenue à partir d'une analyse fréquentiste [69, 104].

- (b) Approche basée sur l'analyse de critère de survie "continu" utilisant un modèle de régression Bayésien (semi-paramétrique ou paramétrique).

Supposant que l'application de l'approche Bayésienne pouvait être différente selon qu'elle est utilisée pour le monitoring de l'essai ou pour l'analyse finale seulement, nous avons donc décidé d'étudier également les caractéristiques des essais selon ces deux types d'analyse.

Les données ont été extraites indépendamment par deux lecteurs. En cas de divergence ou d'incertitude quant à l'interprétation, deux autres lecteurs ont revu l'article concerné pour parvenir à un consensus.

2.3 Résultats

2.3.1 Sélection des articles

À partir de la stratégie de recherche définie, 662 articles ont été identifiés (Figure 2.1) : 488 à partir de la première partie de la stratégie de recherche (au travers des bases de données *PubMed* et *Web of Science*), et 174 à partir de la seconde partie de la recherche (au travers des sites internet des 28 journaux sélectionnés, cités dans l'Annexe B).

Après avoir revu le résumé et/ou le texte intégral de chaque article, nous avons exclu :

- 279 articles rapportant une étude de méthodologie statistique,
- 105 articles utilisant une approche Bayésienne dans un autre but que celui d'analyser un critère de survie,
- 94 articles de méta-analyses ou revues systématiques,
- 63 articles qui mentionnaient les méthodes Bayésiennes sans les utiliser pour leurs analyses,
- 34 articles rapportant une étude non-interventionnelle,
- 30 articles qui n'étaient pas des recherches originales (lettre à l'éditeur, commentaire, revue...),
- 29 articles rapportant une étude biologique ou environnementale.

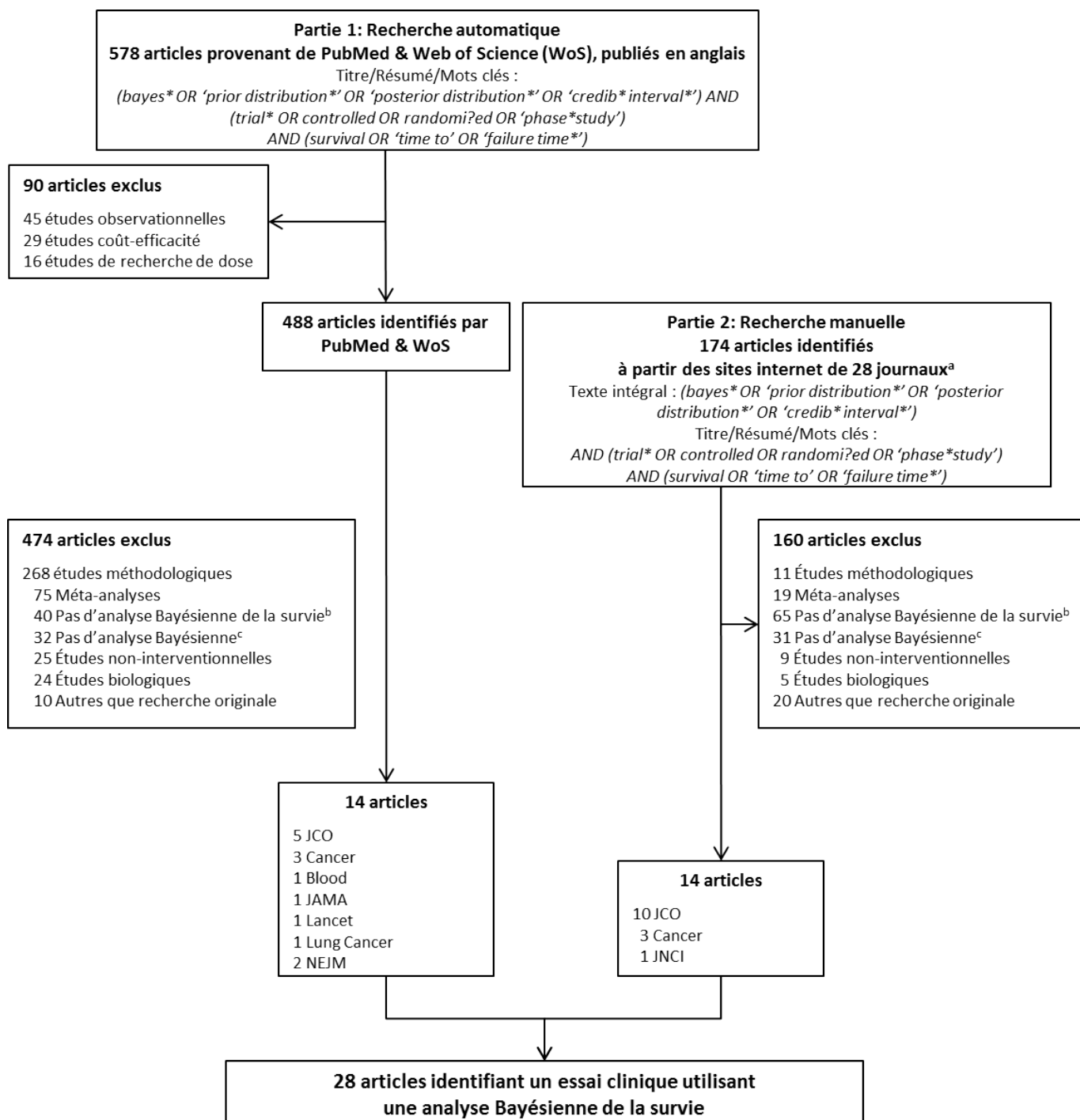


FIGURE 2.1 – Diagramme de flux de la revue systématique de la littérature (sans limite de date)

^a La sélection des 28 journaux est détaillée dans l'Annexe B

^b Ces articles incluait une analyse Bayésienne mais non appliquée à un critère censuré

^c Dans la plupart des cas, l'approche Bayésienne était citée en tant que point de discussion, tandis que l'essai était analysé avec une approche fréquentiste

Aucune divergence n'a été observée entre les deux lecteurs de ces articles. Au total, nous avons identifié 28 articles correspondant aux critères de l'étude, 14 à partir de la recherche automatique et 14 à partir de la recherche manuelle. Les caractéristiques principales de ces articles sont détaillées dans le Tableau D1 de l'Annexe D [111–138].

La recherche complémentaire dans la base d'essais *clinicaltrial.gov* n'a pas permis d'identifier d'autres essais avec une analyse Bayésienne de la survie.

2.3.2 Caractéristiques générales des essais

Les 28 articles ont été publiés entre 1994 et 2014, dont 17 publiés au cours des cinq dernières années. Les articles sélectionnés ont été publiés dans huit journaux différents.

Au total, 25 articles étaient des rapports originaux d'essais cliniques, tandis que les trois articles restants décrivaient la ré-analyse d'un essai précédemment publié.

La majorité des essais sélectionnés étaient en oncologie, même lorsque nous nous sommes restreints aux 14 articles provenant de la recherche automatique via *PubMed* et *Web of Science* (11/14). Nous avons classé comme rare la pathologie étudiée pour 13 essais, mais la rareté de la maladie n'était soulignée par les auteurs que dans sept de ces articles (Tableau 2.1).

La majorité des essais évaluait un traitement médicamenteux (24/28), trois des dispositifs médicaux, et un seul la radiothérapie. L'objectif de dix des essais inclus était de comparer l'efficacité de traitements couramment utilisés, sept essais testaient un traitement pour une nouvelle indication, six évaluaient une nouvelle combinaison de traitements, trois une nouvelle modalité de traitement, tandis que deux essais considéraient un nouveau traitement avant l'autorisation de mise sur le marché.

Caractéristiques générales de l'essai	N(%)
Promoteur et financeur	
Secteur privé seulement	12 (43)
Secteur public seulement	6 (21)
Secteur privé et public	7 (25)
Non spécifié	3 (11)
Pays	
Amérique du Nord	13 (46)
Europe	7 (25)
Asie	2 (7)
Amérique du Nord et Europe	1 (4)
Mondial	2 (7)
Non spécifiée	3 (11)
Localisation de l'essai	
International	11 (39)
Multicentrique national	11 (39)
Multicentrique, localisation non spécifiée	1 (4)
Centre unique	3 (11)
Non spécifiée	2 (7)
Domaine médical	
Oncologie	25 (89)
Cardiologie	2 (7)
Obstétrique	1 (4)
Age des patients	
Adultes uniquement	24 (86)
Enfants, adolescents +/- adultes	3 (11)
Foetus	1 (4)
Maladie rare considérée^a	
Oui	13 (46)
Non	15 (54)
Année de début de recrutement	
Médiane (min-max)	2003 (1973-2009)
Durée de recrutement	
Médiane (min-max)	4 ans (11 mois - 10 ans)
Taille d'échantillon rapportée	
Médiane (min-max)	228 (20-3871)

TABLEAU 2.1 – **Caractéristiques générales des 28 essais identifiés**

^a La prévalence estimée des maladies étudiées par les essais a été extraite du site web Orphanet (<http://www.orpha.net/consor/cgi-bin/index.php>). Dans cette revue, les maladies ont été classées comme rares si la prévalence estimée était inférieure à 5 pour 10 000 personnes en Europe.[1]

2.3.3 Design et conduite des essais

Le Tableau 2.2 montre que la majorité des essais étaient des phases II ou III. Un seul essai était une phase I avec une modélisation Bayésienne de la survie en analyse secondaire. La survie globale était utilisée comme critère de survie dans 22 essais tandis qu'un critère composite (survie sans progression, survie sans événement...) était utilisé dans les six autres essais.

Quatorze essais ont inclus des analyses intermédiaires Bayésiennes de données de survie (monitoring de l'essai). Pour seulement trois de ces 14 essais, l'analyse finale était également réalisée selon une approche Bayésienne, les 11 autres planifiant d'utiliser une approche fréquentiste pour l'analyse finale. L'analyse intermédiaire était non planifiée dans le protocole pour cinq des 14 essais. Au total, neuf essais ont été arrêtés précocement pour futilité ou bénéfice suite à cette analyse intermédiaire.

Quatorze essais utilisaient l'approche Bayésienne pour l'analyse finale mais pas pour les analyses intermédiaires. Cette approche était utilisée pour l'analyse principale d'un essai original dans quatre cas, pour l'analyse secondaire dans sept cas, et pour la ré-analyse d'un essai dans trois cas. Le critère de survie sur lequel se basait l'analyse Bayésienne était la survie globale dans dix essais. L'analyse Bayésienne était réalisée en complément d'une analyse fréquentiste dans dix articles.

Moins de la moitié des essais inclus (11/28) justifiait l'utilisation des statistiques Bayésiennes. Lorsqu'elle l'était, deux raisons principales étaient avancées :

- Surmonter les problèmes inhérents à la petite taille de l'échantillon.
- Quantifier l'incertitude autour de l'effet du traitement.

Caractéristiques de l'essai	N(%)
Phase de l'essai	
Phase I ^a	1 (4)
Phase II	9 (32)
Phase III	13 (46)
Non spécifiée	5 (18)
Aveugle	
Ouvert	25 (89)
Double aveugle	3 (11)
Design de l'essai	
Essai randomisé	21 (75)
Groupes parallèles ^b	18
Plan factoriel	3
Essai non randomisé, simple bras	7 (25)
Design adaptatif planifié^c	
Monitoring pour efficacité	14 (50)
1-3 analyses intermédiaires planifiées	9
4 ou plus analyses intermédiaires planifiées	4
Nombre d'analyses intermédiaires non spécifiées	1
Monitoring pour toxicité	1 (4)
Continual Reassessment Method ^a	1 (4)
Randomisation adaptative	2 (7)
Pas de design adaptatif	10 (36)
Utilisation de l'analyse Bayésienne de la survie	
Pour l'analyse finale seulement	14 (50)
Pour le monitoring de l'essai seulement ^d	11 (39)
Pour le monitoring et l'analyse finale	3 (11)

TABLEAU 2.2 – Caractéristiques du design et de la conduite des 28 essais identifiés

^a Phase I avec une modélisation Bayésienne de la survie en analyse secondaire.

^b Seize essais étaient des groupes parallèles à deux bras et deux essais étaient des groupes parallèles à trois bras. L'un des essais à deux bras avait un design en cross-over où les traitements étaient échangés après progression, le critère de jugement principal étant la survie sans progression.

^c Incluant une approche fréquentiste ou Bayésienne pour le design adaptatif.

^d Monitoring planifié dans six essais et non planifié dans cinq essais.

Une référence était indiquée pour la méthode Bayésienne utilisée dans 21/28 articles ; les six principales références étant Fayers et al. [104] (N = 4, Approche hybride) ; Gelman et al. [139] (N = 3, Analyse Bayésienne pour différents types de modèles) ; Thall & Sung [140] (N = 2, Monitoring Bayésien) ; Ibrahim et al. [110] (N = 2, Analyse de survie Bayésienne) ; et des papiers plus généraux (Berry [58] N = 3 ; et Brophy [141] N = 2).

2.3.4 Modélisation Bayésienne de la survie

Une analyse Bayésienne d'un critère de survie "continu" était décrite dans 24 articles, parmi lesquels 19 étaient des essais contrôlés randomisés et cinq étaient des essais non-comparatifs à simple bras. Deux essais randomisés rapportaient des analyses Bayésiennes de la survie à un temps donné, au sein desquels l'effet du traitement était mesuré en termes d'effet absolu du traitement (différence de taux de survie à un temps donné entre les deux bras de traitement) (Tableau 2.3). Enfin, pour deux essais non randomisés, l'analyse Bayésienne portait sur l'estimation ponctuelle d'un taux de survie.

Concernant les analyses Bayésiennes réalisées dans le cadre d'essais randomisés ayant un critère de survie "continu" (N = 19), l'hypothèse des risques proportionnels était mentionnée dans 17 cas. Trois articles utilisaient un modèle Bayésien de survie semi-paramétrique, mais le processus utilisé pour représenter le risque de base n'était spécifié que dans un seul cas qui utilisait un processus exponentiel par morceaux. Parmi les neuf articles utilisant un modèle de régression paramétrique ou semi-paramétrique pour un essai randomisé, cinq ajustaient leur modèle sur des covariables et deux d'entre eux incorporent une interaction entre le traitement et une covariable.

Parmi les neuf articles rapportant un essai randomisé ou non randomisé avec un modèle de survie Bayésien paramétrique (modèle exponentiel à un paramètre, modèle log-normal, et modèle paramétrique non détaillé), la sélection du modèle était justifiée dans cinq cas, en utilisant le *Bayesian Information Criterion* pour évaluer la qualité de l'ajustement du modèle [142].

Nous avons fait l'hypothèse que la modélisation Bayésienne pouvait être différente selon qu'elle est utilisée pour le monitoring de l'essai ou pour l'analyse finale seulement. Après analyse des articles retenus, parmi les 14 essais utilisant l'analyse de survie Bayésienne pour le suivi intermédiaire des données d'efficacité, seulement un essai implémentait un modèle Bayésien semi-paramétrique, tandis qu'une approche plus simple était utilisée dans les 13 autres essais (analyse de survie à un temps donné, approche hybride, modèle exponentiel à un paramètre). Parmi les 14 essais utilisant l'approche Bayésienne pour l'analyse finale uniquement, sept ont utilisé un modèle de régression Bayésien.

	Analyse Bayésienne utilisée pour		Total
	Analyses intermédiaires	Analyse finale seulement	N (%)
	N = 14 ^a	N = 14	
Modélisation Bayésienne dans les essais randomisés (N = 21)	10	11	21
Survie à un temps donné	1	1	2 (10)
Approche hybride	5	5	10 (48)
Modèle semi-paramétrique	1	2	3 (14)
Modèle exponentiel à un paramètre	3	-	3 (14)
Modèle Log-normal	-	2	2 (10)
Modèle paramétrique non détaillé	-	1	1 (5)
Modélisation Bayésienne dans les essais non-randomisés (N = 7)	4	3	7
Survie à un temps donné	2	-	2 (29)
Modèle exponentiel à un paramètre	1	-	1 (14)
Modèle Log-normal	-	2	2 (29)
Modélisation d'un critère de survie "continu" non détaillé	1	1	2 (29)
Prior de l'effet du traitement (essais randomisés)	10	11	21
Données historiques + opinions d'experts + non-informatif	-	1	1 (5)
Données historiques + priors archétypaux + non-informatif	1	-	1 (5)
Données historiques seulement	-	2	2 (10)
Priors archétypaux + non-informatif	1	-	1 (5)
Non-informatif seulement	4	6	10 (48)
Non spécifiée	4	2	6 (29)
Prior des paramètres du modèle (essais non-randomisés)	4	3	7
Non-informatif	-	2	2 (29)
Non spécifiée	4	1	5 (71)
Effet minimum attendu pré-spécifié			
Oui	11	5	16 (57)
Non	3	9	12 (43)
Règle de décision pré-spécifiée			
Oui	8	1	9 (32)
Non	6	13	19 (68)

TABLEAU 2.3 – Spécification de l'analyse de survie Bayésienne, globalement et selon son objectif (analyses intermédiaires ou analyse finale) pour les 28 essais identifiés

Les valeurs en gras représentent les totaux et sous-totaux pour chaque catégorie principale

^a Parmi les 14 essais réalisant une analyse Bayésienne de la survie pour le monitoring de l'essai, trois avaient aussi conduit une analyse finale Bayésienne de la survie

2.3.5 Choix de la distribution a priori

En ce qui concerne les 21 essais randomisés, seule une distribution a priori non informative était utilisée pour l'effet du traitement dans l'analyse de 11 essais (complétée par des distributions a priori archétypales pour un essai).

Parmi les 10 autres essais, quatre incorporaient des données historiques dans la distribution a priori de l'effet du traitement. Trois de ces essais fournissaient les références bibliographiques des essais historiques utilisés, et un seul rapportait le nombre de patients inclus dans l'essai historique. La méthode pour incorporer les données historiques dans l'essai actuel était précisée dans trois de ces articles : ils regroupaient sans pondération les données historiques et actuelles. Deux des quatre essais incorporant des données historiques réalisaient des analyses pour évaluer la sensibilité des résultats au choix de la distribution a priori. Seul un article utilisait, en plus des données historiques, les opinions d'experts pour formuler la distribution a priori de l'effet du traitement, basée sur l'agrégation mathématique de l'opinion de 17 experts.

Six articles ne précisaient pas la distribution a priori utilisée pour l'effet du traitement.

Aucun des neuf articles rapportant un modèle de régression n'utilisait une distribution informative sur l'effet du traitement (distribution a priori non-informative pour sept essais ; et non spécifiée pour deux articles).

Pour 16 des 19 essais randomisés analysant un critère de survie "continu", l'estimation Bayésienne portait sur l'effet relatif du traitement (HR). Parmi eux, sept utilisaient une distribution normale pour représenter l'information a priori sur l'effet du traitement. Cela contribuait à une analyse hybride dans quatre essais, à une analyse paramétrique basée sur un modèle de survie log-normal dans un essai, à une analyse paramétrique non détaillée dans un essai, et à une analyse semi-paramétrique supposant un processus exponentiel par morceau pour le risque de base dans un essai. Ce dernier article plaçait une distribution gamma non-informative sur la distribution a priori des paramètres du risque de base. Neuf essais (six utilisant une approche hybride, deux utilisant un modèle semi-paramétrique, et

un utilisant un modèle paramétrique) ne précisaient pas la distribution a priori utilisée.

Au total, trois des 19 essais randomisés ont utilisé des modèles exponentiels à un paramètre pour les données du groupe contrôle et du groupe expérimental; un essai plaçait une distribution gamma sur la distribution a priori du risque de base dans chaque groupe de traitement tandis que les deux essais restants ne précisaient pas la distribution utilisée.

Parmi les cinq essais non-randomisés analysant un critère de survie "continu", deux définissaient une distribution normale pour la distribution a priori de l'effet des covariables du modèle de survie tandis que les trois autres essais ne rapportaient pas la distribution a priori utilisée pour les paramètres du modèle.

Trois des quatre essais intégrant des données historiques utilisaient l'approche Bayésienne pour l'analyse finale. La plupart des essais utilisant l'approche Bayésienne pour le monitoring ne spécifiaient pas la distribution a priori utilisée pour le paramètre d'intérêt : seuls deux essais utilisaient des distributions a priori archétypales, telles que recommandé par Parmar et al. [64], l'un de ces essais utilisant également des données historiques.

2.3.6 Règles de décision

La majorité des essais utilisant une approche Bayésienne pour le monitoring pré-spécifiaient un effet minimum attendu (11/14) et une règle de décision (8/14), tandis qu'un seul des 14 essais utilisant une approche Bayésienne pour l'analyse finale définissait la règle de décision Bayésienne à l'étape du design de l'essai (Tableau 2.3). La plupart des essais utilisant les méthodes Bayésiennes pour le monitoring cadraient leurs règles de décision en termes de probabilités prédites (Voir Tableau D1 Annexe D). À noter qu'au moment de l'analyse finale, aucun de ces essais n'a pris en compte le fait d'avoir réalisé des analyses intermédiaires. Onze essais sur 14 présentant une analyse finale Bayésienne, ont utilisé une règle de décision équivalente à la règle fréquentiste : $P(HR < 1)$ ou $P(\text{différence entre les traitement}) > 97,5\%$.

2.3.7 Qualité et exhaustivité des résultats de l'analyse Bayésienne de la survie

Parmi les 25 articles décrivant des analyses Bayésiennes dans le rapport original de l'essai, sept ne mentionnaient pas l'analyse de survie Bayésienne dans la section "Méthodes" du papier : cinq d'entre eux utilisaient les méthodes Bayésiennes pour le monitoring de l'essai, et deux les utilisaient pour l'analyse finale secondaire de l'essai. Pour 18 articles, les résultats des analyses Bayésiennes n'étaient pas mentionnés dans le résumé (*Abstract*) du papier.

Comme détaillé dans le Tableau 2.4, la plupart des articles ne donnaient aucune information sur la distribution a priori, et la description complète de la distribution n'était donnée que dans sept articles. En général, les distributions a posteriori étaient mieux rapportées, avec quelques indications sur ces distributions données dans 25 articles. Cependant, moins de la moitié rapportait une mesure de position de la distribution a posteriori (moyenne, variance), et son intervalle de crédibilité associé.

La plupart des articles (19/28) ne précisaient pas le logiciel utilisé pour la réalisation des analyses Bayésiennes. Les logiciels WinBUGS ou OpenBUGS étaient utilisés dans cinq cas (toujours dans le but de réaliser un modèle de régression Bayésien), tandis que le logiciel SAS était utilisé dans deux cas seulement, bien que l'analyse de survie Bayésienne soit possible dans SAS depuis 2009. Le logiciel Fortran était utilisé dans un essai, tout comme le logiciel BUGS.

Présentation des distributions des paramètres dans les articles	N(%)
Distribution a priori	
Définition mathématique de la distribution	7 (25)
Représentation graphique	2 (7)
Mesure de position	8 (29)
Intervalle de crédibilité	1 (4)
Non rapportée	20 (71)
Distribution a posteriori	
Définition mathématique de la distribution	3 (11)
Représentation graphique	3 (11)
Intervalle de crédibilité/Intervalle HPD ^b	10 (36)
Probabilités a posteriori/prédites	24 (86)
Mesure de position	12 (43)
Non rapportée	3 (11)

TABLEAU 2.4 – **Rapport et caractérisation des distributions a priori et a posteriori^a**

^a Le total n'est pas de 28 car les différents items de ce tableau ne sont pas exclusifs

^b HPD : Highest Posterior Density

Suivi des recommandations

Des recommandations pour le rapport d'analyses Bayésiennes existent (BayesWatch [66] et ROBUST [143]). Elles indiquent notamment que les articles devraient préciser la distribution a priori, le modèle statistique, les techniques de calcul utilisées, ainsi que les distributions a posteriori qui en découlent.

Dans notre revue, trois articles spécifiaient, justifiaient, et rapportaient des analyses de sensibilité de la distribution a priori choisie. Les détails du modèle de survie (justification / définition du modèle / vérification du modèle) étaient fréquemment manquants. Trois essais ne définissaient pas le modèle de survie utilisé (un modèle paramétrique sans détails supplémentaires; deux complètement non spécifiés). Par ailleurs, deux des trois essais utilisant un modèle Bayésien semi-paramétrique ne précisaient pas si un processus stochastique était utilisé pour modéliser le risque de base. En matière de sélection du

modèle, cinq des neuf essais utilisant un modèle de survie paramétrique décrivaient la manière par laquelle le modèle final avait été sélectionné.

Seuls quatre essais mentionnaient l'utilisation de l'algorithme de *Markov Chain Monte Carlo* pour les inférences a posteriori, et deux de ces essais spécifiaient le nombre de chaînes et la longueur du *burn-in*.

Parmi les 17 essais utilisant les méthodes Bayésiennes pour l'analyse finale (analyse finale seule, N = 14, ou monitoring avec analyse finale Bayésienne, N = 3), 12 rapportaient une mesure de position ou l'intervalle de crédibilité pour la distribution a posteriori.

2.3.8 Maladies rares

Parmi les articles inclus, 13 rapportaient un essai portant sur une maladie rare. Ces essais étaient majoritairement des essais de phase II (8/13) ; sept étaient des essais randomisés.

Les méthodes Bayésiennes étaient utilisées pour l'analyse finale de huit essais (analyse principale, N = 2 ; analyse secondaire, N = 6), et pour le monitoring de cinq essais.

Une analyse Bayésienne d'un critère de survie "continu" était décrite dans dix articles, tandis que trois articles décrivaient une analyse de survie Bayésienne à un temps donné.

Parmi les dix articles rapportant une analyse Bayésienne d'un critère "continu", sept décrivaient une modélisation Bayésienne de la survie au travers d'un modèle paramétrique (N = 6) ou semi-paramétrique (N = 1). Les trois articles restants décrivaient une approche hybride (N = 2), ou ne précisaient pas le modèle utilisé (N = 1).

La majorité des essais utilisaient une distribution a priori non-informative (complétée ou non par des distributions a priori archétypales) pour le paramètre d'intérêt (N = 8). Un seul essai utilisait des données historiques pour construire sa distribution a priori, et quatre essais ne spécifiaient pas la distribution utilisée.

2.4 Discussion

L'étape de sélection des articles nous a permis de confirmer le grand nombre de publications rapportant des développements méthodologiques, comme le note Ashby dans sa revue [144]. Contrastant avec cette littérature abondante, peu d'essais ont implémenté les méthodes Bayésiennes lorsqu'un critère censuré était considéré, et peu d'entre eux ont utilisé une approche Bayésienne pour leur analyse principale. Il est impressionnant de noter que seuls 28 articles ont finalement été retenus comme correspondant à des essais avec une analyse Bayésienne de la survie sur 488 articles pré-sélectionnés, la majeure partie des articles écartés correspondant à des travaux méthodologiques.

Lors de cette revue, il a été mis en évidence qu'un certain nombre d'essais (11 des 28 essais, soit 39%) utilisaient les méthodes Bayésiennes pour le monitoring de l'essai mais pas pour l'analyse finale de l'essai. Il est possible que l'attraction envers le Bayésien pour le monitoring de l'essai vienne de sa plus grande flexibilité et de son interprétation plus aisée dans le processus de prise de décision comparée à l'approche fréquentiste.

Dans le contexte des pathologies rares, de nombreux articles ont exposé les bénéfices potentiels des méthodes Bayésiennes pour l'interprétation des essais cliniques, en particulier en augmentant la puissance de ces essais par l'intégration de données historiques pertinentes [29, 36–39]. Cette revue de la littérature n'a cependant mis en évidence que 13 essais portant sur des pathologies rares et conduisant des analyses Bayésiennes de la survie et, parmi eux, un seul essai incorporait des données historiques dans la distribution a priori de l'effet du traitement.

L'utilisation encore limitée de l'approche Bayésienne (hors essais de phase I) identifiés par cette revue est cohérente avec les recherches de Pibouleau & Chevret [145] qui ont évalué l'utilisation des méthodes Bayésiennes dans les essais portant sur les dispositifs médicaux et ont mis en évidence 12 études cliniques. La revue de la littérature réalisée par Chevret [146] sur les essais adaptatifs entre 1982 et 2010 montre que, malgré les recommandations et les nombreux développements méthodologiques, les designs adaptatifs Bayésiens

étaient peu utilisés dans la pratique. Une revue d'essais conduits au *MD Anderson Cancer Center*, connu pour son intérêt particulier pour l'approche Bayésienne, a montré que les méthodes Bayésiennes tendaient à être plus utilisées dans les phases précoces mais étaient rarement utilisées pour les phases II ou III où un critère censuré est plus souvent mesuré [45].

Il serait intéressant d'étendre notre revue de la littérature à d'autres types de critères de jugement afin de mettre en lumière les potentielles spécificités des essais avec un critère de jugement censuré. Cependant, dans les revues de la littérature déjà publiées portant sur l'analyse Bayésienne d'essais cliniques, à savoir les essais réalisés au *MD Anderson Cancer Center* [45], les essais adaptatifs [146], les essais portant sur les dispositifs médicaux [145], la question du choix du critère de jugement et de l'impact de ce choix sur le type d'analyse n'est pas discutée.

Dans sa revue de la littérature, Chevret a observé une multiplication des articles méthodologiques rapportant des designs adaptatifs Bayésiens à partir de 1992, avec une implémentation plus lente au cours du temps de ces méthodes dans les essais cliniques réels. Dans notre étude, seuls cinq essais Bayésiens avaient été initiés entre 1973 et 1999, 10 entre 2000 et 2004, et neuf entre 2005 et 2009 (quatre articles, publiés entre 2002 et 2010, ne mentionnaient pas leur date de début de recrutement). Même si cette estimation n'est peut être pas fiable compte tenu du délai possible de publication des articles, en particulier quand il s'agit de critère de survie, nous observons donc également une tendance en faveur d'une plus grande utilisation de l'approche Bayésienne dans les essais avec critère censuré. Par ailleurs, parmi les 28 articles identifiés, nous n'avons pas observé de tendance claire quant à l'implémentation de méthodes plus sophistiquées au cours du temps. Cependant le petit nombre d'articles nous oblige à une certaine prudence dans l'interprétation de cette observation.

De nombreuses méthodes ont été développées pour incorporer des données historiques dans la distribution a priori d'un paramètre d'intérêt [43, 78, 82, 84, 107], mais elles sont peu utilisées en pratique et, malgré leurs avantages, les auteurs semblent leur préférer les

distributions a priori non-informatives.

Nous sommes conscients que cette revue comporte certaines limites. Seule la moitié des articles inclus a été trouvée via nos recherches au sein de *PubMed* et *Web of Science*. Aucun article n'a été identifié dans les 23 journaux hors cancérologie explorés directement par la recherche dans le texte des articles. Nous avons essayé d'étendre nos recherches en utilisant le registre de *clinicaltrials.gov* mais n'avons pas trouvé d'essais supplémentaires. Cependant, malgré nos efforts pour être exhaustifs, il est probable que certains essais nous aient échappé car l'analyse Bayésienne de la survie n'était pas mentionnée explicitement comme telle dans le titre, le résumé, ou les mots-clés de l'article. Il est aussi possible que certains articles aient été omis car l'analyse de survie était référencée selon d'autres mots clés que ceux utilisés dans notre stratégie de recherche. Pour pallier cette difficulté, nous avons complété notre recherche automatique par une recherche manuelle dans le corps intégral des articles. Cette recherche manuelle a été fructueuse, puisqu'elle a permis d'identifier 14 articles supplémentaires aux 14 identifiés par la recherche automatique. À noter que 10 des 15 articles identifiés dans *Journal of Clinical Oncology*, ont échappé à la recherche automatique (via *PubMed* et *Web of Science*). Pour huit des 10 articles concernés, ceci s'explique par le fait que l'analyse bayésienne n'était pas l'analyse principale présentée dans le résumé de l'article. Cependant, de façon surprenante, pour deux de ces articles, l'analyse Bayésienne correspondait pourtant à l'analyse principale.

Il faut reconnaître que notre recherche manuelle sur le texte intégral des journaux n'a pu être exhaustive car elle se concentrait sur une sélection de journaux dont le contenu des articles était interrogeable en ligne, avec une sur-représentation des journaux en oncologie. Ainsi, comme détaillé en Annexe B, il n'est pas possible d'interroger les journaux tels que le *Lancet*, et le *JAMA* par exemple. La décision d'explorer plus de journaux en oncologie que dans les autres spécialités a été prise car 11 des 14 articles identifiés par la recherche automatique (via *PubMed* et *Web of Science*) décrivaient des essais en oncologie.

Des articles dont l'accès au texte intégral n'était pas possible ont pu être omis par cette recherche manuelle tels que l'article princeps rapportant l'essai CHART, publié dans

Radiotherapy and Oncology [147], dont les analyses intermédiaires ont été réalisées en utilisant une approche de monitoring Bayésien. De plus, notre recherche manuelle, journal par journal, n'a porté que sur une sélection de journaux à fort facteur d'impact pour des raisons de faisabilité. Cette limite réduit de fait la quantité d'information disponible et on ne peut exclure un biais.

Concernant les critères d'exclusion, il est également possible que certains articles décrivant des essais cliniques aient pu mentionner les termes *epidemiology* ou *case-control* dans l'introduction de leur résumé et auraient donc été exclus de notre revue à tort. Cependant, nous avons vérifié tous les articles satisfaisant nos critères d'exclusion et n'avons pas trouvé de tels articles.

Le choix d'exclure les articles présentant une analyse Bayésienne de la survie d'un essai dans le seul but d'illustrer une méthode statistique peut être également jugée regrettable. Ce choix était justifié par l'objectif de notre revue qui était d'évaluer l'implémentation de ces méthodes dans la pratique.

En dépit de ces limites, nous pensons que cette revue fournit une bonne vue d'ensemble des méthodes utilisées en pratique pour suivre et analyser les essais cliniques dans un contexte Bayésien lorsqu'un critère censuré est considéré.

— Chapitre 3 —

Proposition d'un design Bayésien avec incorporation de données historiques individuelles et agrégées pour un essai clinique

Lorsqu'une approche classique fréquentiste est utilisée, on ignore dans l'expression des résultats toutes les données historiques extérieures à l'essai. L'approche Bayésienne peut être vue comme un complément intéressant à l'approche classique fréquentiste, car elle permet d'intégrer, dans l'analyse d'un essai clinique, des données extérieures à l'essai [66, 148–151]. Il est ainsi possible de faire une synthèse chiffrée des données historiques que les cliniciens ont en tête. Différentes méthodes existent pour incorporer des données historiques dans la distribution a priori et les pondérer [43, 82–84, 88, 90, 91, 152, 153] (voir Chapitre 1, Section 1.2.3). Cependant, ces approches sont peu implémentées en pratique comme l'a démontré notre revue de la littérature (Chapitre 2).

L'intégration de données historiques est en effet complexe et ouverte à la critique. Cette complexité augmente lorsque l'on souhaite prendre en compte à la fois des données historiques individuelles sur le bras contrôle d'un nouvel essai et des données historiques agrégées, obtenues à partir d'une revue de la littérature, sur l'effet du traitement.

Lors de la planification d'un essai clinique portant sur une pathologie rare, des obstacles méthodologiques s'ajoutent car il y aura peu de données préexistantes permettant de guider les investigateurs, d'une part sur le poids à assigner aux données historiques, et d'autre part sur le modèle de survie le plus approprié pour modéliser le risque de base. L'impact de l'intégration de données individuelles historiques sur le bras contrôle sur l'estimation de l'effet du traitement reste encore à évaluer.

Motivé par l'essai Sarcome-13, un essai clinique contrôlé randomisé alors en cours de développement et évaluant un traitement pour les patients atteints d'ostéosarcome à haut risque de rechute, nous proposons un design Bayésien basé sur une approche utilisant un *power prior*, développé par Ibrahim et al. [43], et un *mixture prior*, tel que présenté par Mutsvari et al. [91], pour incorporer des informations dérivées de données historiques individuelles et agrégées. Les caractéristiques opératoires de l'analyse Bayésienne de la survie basées sur ces distributions a priori ont été évaluées par une étude de simulations pour différents scénarios représentant différents degrés de commensurabilité entre les données historiques et actuelles.

3.1 Essai Sarcome-13

L'ostéosarcome est une tumeur maligne des os qui représente moins de 10% de tous les cancers de l'adolescent et du jeune adult. On compte entre 0.2 et 3 cas d'ostéosarcome par million et par an en Europe. Les patients présentant des métastases ou ayant une mauvaise réponse histologique à la chimiothérapie néoadjuvante ont, aujourd'hui encore, un pronostic sombre.

3.1.1 Évaluations pré-existantes du mifamurtide dans l'ostéosarcome

L'intérêt pour le mifamurtide dans le traitement de l'ostéosarcome provient des résultats remarquables obtenus chez le chien, pour lequel l'ostéosarcome spontané est commun et très similaire à l'ostéosarcome chez l'humain. Des études randomisées chez des chiens traités par mifamurtide seul ou en combinaison avec une chimiothérapie ont montré une amélioration significative de la survie chez les chiens traités par mifamurtide [154, 155].

L'essai de phase III INT-0133 évaluait, à l'aide d'un plan factoriel, le bénéfice associé à l'addition de mifamurtide d'une part et d'ifosfamide d'autre part à la chimiothérapie standard pour le traitement des ostéosarcomes chez les patients âgés de moins de 31 ans. Bien que l'essai ait porté sur l'ensemble de la population, l'analyse a été publiée séparément en deux strates distinctes : les ostéosarcomes localisés [156, 157], et les ostéosarcomes métastatiques [158]. L'effet du mifamurtide sur la survie globale, ainsi que sur la survie sans événement (EFS), était homogène entre les deux strates, même si le bénéfice du mifamurtide pour la survie globale était significatif chez les ostéosarcomes localisés, mais non-significatif chez les ostéosarcomes métastatiques.

De plus, une possible interaction entre l'addition d'ifosfamide et de mifamurtide, rapportée par Meyers et al. dans leur première publication [156] mais non-significative dans l'analyse suivante, a compliqué l'interprétation des résultats qui ont suivi [157, 158].

À partir de ces résultats, l'EMA a autorisé la mise sur le marché du mifamurtide le 6 mars 2009. Cependant, cette autorisation européenne ne concerne pas les patients atteints d'une maladie métastatique. Aucun consensus clair n'a été établi suite à l'essai INT-0133 parmi les experts du domaine [159]. De plus, aucune autorisation n'a été délivrée aux États-Unis, et le traitement n'est pas remboursé en France. La commission de transparence française a demandé de plus amples investigations avant que le mifamurtide ne soit approuvé et remboursé pour une utilisation standard en première ligne du traitement des patients atteints d'ostéosarcome.

3.1.2 Mise en place de l'essai Sarcome-13

Dans ce contexte, un nouvel essai contrôlé randomisé (essai Sarcome-13) sera conduit en France afin d'évaluer le mifamurtide chez les patients atteints d'un ostéosarcome à haut risque de rechute (métastatique ou ayant une mauvaise réponse histologique à la chimiothérapie néoadjuvante). L'objectif de cet essai est d'évaluer le bénéfice en termes d'EFS du mifamurtide en combinaison avec la chimiothérapie post-opératoire comparé à la chimiothérapie seule chez les patients de moins de 31 ans. Si l'interaction possible détectée par l'essai INT-0133 [156] entre l'ifosfamide et le mifamurtide existe, elle ne devrait pas nuire à l'essai Sarcome-13 puisque l'ifosfamide fait partie de la chimiothérapie standard délivrée à tous les patients de l'essai.

L'essai a été planifié comme un essai de Phase II randomisé en deux groupes parallèles 1 : 1, en post opératoire entre chimiothérapie seule et chimiothérapie + mifamurtide. Étant face à une pathologie rare, nous avons relâché l'erreur alpha du test unilatéral du log-rank à 10% pour l'approche fréquentiste [32, 151], et une cible de recrutement pragmatique a été fixée à 105 patients sur 3 ans (avec 2 ans de suivi pour le dernier patient inclus).

Si l'objectif de recrutement est atteint, la puissance sera de 80% si le HR est de 0.55 (basé sur 43 événements), ce qui représente un effet du traitement très optimiste. La puissance chute pour un effet moindre du traitement tel que l'effet observé dans l'essai INT-0133 (effet poolé, cf. Section 3.1.4) : pour un HR de 0.786 (48 événements), qui reste un effet cliniquement pertinent, la puissance est de 33%. Pour un HR moitié moins important que celui observé dans l'essai INT-0133 (HR = 0.886 ; 50 événements) la puissance chute à 20%. L'analyse principale de l'essai sera fréquentiste, basée sur un modèle de Cox à risques proportionnels.

À l'aide d'une analyse Bayésienne, nous cherchons à enrichir les données de l'essai avec des informations historiques pertinentes afin d'augmenter sa puissance pour être en mesure de conclure positivement en cas d'effet bénéfique modéré, plus plausible. Lors de cette analyse Bayésienne, le mifamurtide associé à la chimiothérapie post-opératoire sera jugé supérieur à la chimiothérapie post-opératoire seule si la probabilité a posteriori d'un HR inférieur à 1 est supérieure à 0.9 ($P(HR < 1) > 0.9$), ce qui est équivalent à la règle fréquentiste définie dans le protocole de l'essai.

3.1.3 Données historiques individuelles

La première source de données historiques disponible est représentée par les données individuelles de l'essai OS2006 (NCT00470223) traitant des patients atteints d'un ostéosarcome [160]. En effet, l'essai Sarcome-13 sera réalisé dans les mêmes centres participants, incluant des patients enfants, adolescents ou adultes, et la chimiothérapie qui sera administrée à tous les patients de l'essai Sarcome-13 est la même que la chimiothérapie utilisée dans l'essai OS2006.

L'essai OS2006 évaluait l'apport de l'acide zolédronique en association à la chimiothérapie de référence. Au total, 318 patients ont été inclus. Au terme de l'essai, il a été conclu à l'absence de bénéfice associé à l'acide zolédronique.

Nous avons sélectionné, parmi les patients de l'essai randomisé OS2006, tous ceux qui remplissaient les critères d'éligibilité planifiés de l'essai Sarcome-13 : patients de moins de 31 ans au diagnostic, atteints d'un ostéosarcome métastatique ou qui avaient une mauvaise réponse à la chimiothérapie néoadjuvante, qui n'avaient pas eu d'événement lors de la chimiothérapie pré-opératoire, et qui avaient subi une chirurgie de la tumeur primaire. Au total, 165 patients ont été identifiés. Ce sous-groupe de 165 patients (et 73 événements) de l'essai OS2006 sera ensuite référencé en tant que SARC-OS.

Dans le but de correspondre à la durée de l'essai Sarcome-13, nous avons tronqué les données à 5 ans. La courbe de Kaplan-Meier correspondant à ces patients est présentée en Figure 3.1.

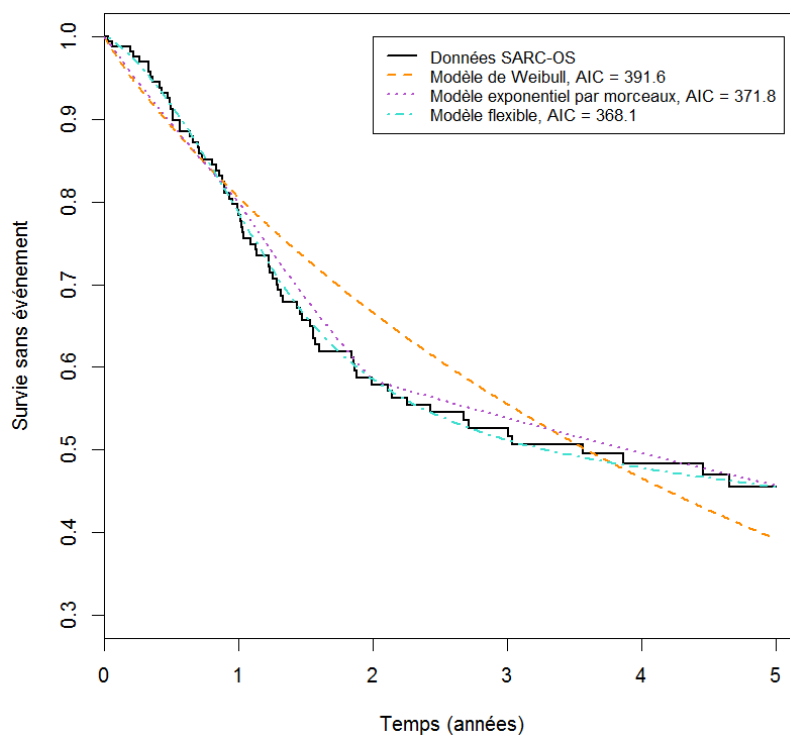


FIGURE 3.1 – Survie sans événement dans le sous-groupe de patients de l'essai historique OS2006 (Données SARC-OS)

Estimation de la courbe de Kaplan-Meier et estimations paramétriques (modèle de Weibull, modèle exponentiel par morceaux à 3 paramètres et modèle flexible de Royston & Parmar) de la survie sans événements pour le sous-groupe de patients de l'essai OS2006 satisfaisant les critères d'éligibilité de l'essai Sarcome-13 (N = 165, 73 événements).

Dans la perspective d'intégrer ces données historiques à l'analyse Bayésienne du futur essai, nous avons fait le choix d'une modélisation paramétrique de la fonction de survie. Nous avons donc cherché le modèle de survie paramétrique qui ajustait au mieux aux données. En limitant notre attention sur les modèles compatibles avec l'hypothèse des risques proportionnels, nous avons comparé l'ajustement d'un modèle de Weibull à celui d'un modèle exponentiel par morceaux à trois paramètres (cité "modèle exponentiel par morceaux" dans le suite du manuscrit) avec différents temps de coupure. Le modèle flexible de Royston & Parmar [161] avec trois degrés de liberté a aussi été modélisé et pris en tant que modèle de référence. Nous avons fait le choix de ne pas retenir ce modèle pour l'analyse Bayésienne de l'essai futur pour des raisons de simplicité et en raison de la petite

taille des échantillons de SARC-OS et de l'essai futur.

En se basant sur le critère d'Akaike [162] (AIC), nous avons sélectionné le modèle exponentiel par morceaux avec des partitions fixées à $[0,1[$, $[1,2[$ et $[2,+\infty[$ années avec un AIC = 371.8. Ce modèle est le plus proche du modèle de référence (AIC = 368.1).

3.1.4 Données historiques agrégées

Comme mentionné précédemment, la seconde source de données historiques disponible comprend les deux estimations de l'effet relatif du traitement, quantifiés par un rapport des risques ou *hazard ratio* (HR) en anglais, sur l'EFS du mifamurtide associé à la chimiothérapie post-opératoire comparé à la chimiothérapie post-opératoire seule, rapportés par l'essai INT-0133 au sein de deux publications séparées (une pour les ostéosarcomes localisés [157] et une pour les ostéosarcomes métastatiques [158]). Nous n'avons pas eu accès aux données individuelles de ces publications et avons donc travaillé à partir des résultats publiés, ce que nous appellerons ensuite "données agrégées".

Étant donné les HRs publiés et leur intervalle de confiance à 95% pour la survie sans événement obtenus à partir d'un modèle de Cox à risques proportionnels chez les patients atteints d'un ostéosarcome localisé (HR = 0.80 ; IC95%[0.62-1.00] ; N = 662) [157], et chez les patients atteints d'un ostéosarcome métastatique (HR = 0.72 ; IC95%[0.42-1.20] ; N = 91) [158], nous avons conclu à l'absence d'hétérogénéité significative de l'effet du traitement entre les deux strates (Q de Cochrane, $p = 0.72$; Index de Higgins, $I^2 = 0$).

Par conséquent, une méta-analyse à effet fixe avec une pondération par l'inverse de la variance a été utilisée pour obtenir une estimation globale de l'effet du traitement (HR = 0.786 ; IC95%[0.63-0.98]). L'estimation correspondante du logarithme du HR ($\hat{\beta}_H$), qui est le paramètre que nous considérerons par la suite, et sa variance (s^2), sont 0.241 et 0.012 respectivement. Cette variance est approximativement équivalente à ce qu'elle serait si l'estimation était basée sur 329 événements (selon la formule de Schoenfeld [163] $s^2 = 4/d$ où d est le nombre d'événements).

3.2 Construction des distributions a priori

Les distributions a priori portent à la fois sur les paramètres du modèle de survie utilisé et sur l'effet du traitement. Dans le cadre de l'essai Sarcome-13, un *power prior* a été choisi pour intégrer les données individuelles issues de SARC-OS, contribuant à la modélisation des données de survie du bras contrôle, et un *mixture prior* pour intégrer les données agrégées issues de l'essai randomisé INT-0133 contribuant à l'estimation de l'effet du traitement.

3.2.1 Power prior : Données individuelles sur le bras contrôle

Le *power prior* a été développé par Ibrahim et al.[43] et consiste à élever la vraisemblance des données historiques à la puissance α_0 , représentant le degré de commensurabilité entre les données historiques et les données du nouvel essai ; α_0 contrôle donc l'impact des données historiques sur la distribution a posteriori.

Deux versions de cette distribution a priori existent selon que le paramètre α_0 est considéré comme fixe ou aléatoire. Lorsqu'il est considéré comme fixe, α_0 peut être déterminé par des experts qui statueront sur la commensurabilité entre les données historiques et les données du nouvel essai, ou encore par une étude de simulations.

Nous avons choisi d'utiliser cette version car elle est plus facile à communiquer aux

3.2. Construction des distributions a priori

cliniciens et ne réagit pas aux différences observées entre les données historiques et actuelles, sachant que distinguer une vraie hétérogénéité entre l'essai historique et l'essai actuel d'une hétérogénéité aléatoire s'avère difficile voire impossible dans le contexte des pathologies rares.

Soit θ le vecteur de paramètres du modèle de survie choisi pour représenter l'EFS du bras contrôle, et $\pi_0(\theta)$ représentant la distribution a priori initiale, c'est-à-dire la distribution a priori de θ avant que les données individuelles historiques contrôles D_C^H ne soient observées (où $\pi_0(\theta)$ peut par exemple, être le produit de distributions a priori non-informatives indépendantes pour chaque élément de θ). Dans notre cas, D_C^H représente les données individuelles des patients SARC-OS (voir la Section 3.3.1). Nous définissons donc la distribution du *power prior* de θ comme :

$$\pi(\theta|D_C^H, \alpha_0) \propto L(\theta|D_C^H)^{\alpha_0} \pi_0(\theta) \quad (3.1)$$

où α_0 est une constante fixée telle que $0 \leq \alpha_0 \leq 1$.

Lorsque $\alpha_0 = 0$, $\pi(\theta|D_C^H, \alpha_0) \equiv \pi_0(\theta)$, ce qui signifie que les données historiques ne sont pas intégrées dans la distribution a priori.

Lorsque $\alpha_0 = 1$, un poids égal est donné à la vraisemblance des données historiques $L(\theta|D_C^H)$ et à la vraisemblance des données du nouvel essai $L(\theta|D_C^N)$ dans la distribution a posteriori qui est définie par :

$$\pi(\theta|D_C^N, D_C^H, \alpha_0) \propto L(\theta|D_C^N) \pi(\theta|D_C^H, \alpha_0) \quad (3.2)$$

où D_C^N représente les données individuelles des patients du nouvel essai.

La dimension de θ dépendra de la spécification du modèle de survie : pour un modèle de Weibull, $\theta = \theta^W = (\beta_0, \gamma)$ comprenant l'*intercept* (β_0) et le *scale* (γ); et pour un modèle exponentiel par morceaux, $\theta = \theta^P = (\lambda_1, \lambda_2, \lambda_3)$ où λ_j représente la fonction de risque instantané du modèle exponentiel pour le $j^{\text{ème}}$ intervalle de temps.

La fonction de vraisemblance des données historiques pour n patients dans le contexte d'un modèle de Weibull est définie comme :

$$L(\beta_0, \gamma | D) = \prod_{i=1}^n \left[\exp\{\beta_0\} \frac{1}{\gamma} t_i^{\frac{1}{\gamma}-1} \right]^{\nu_i} \exp \left\{ -\exp\{\beta_0\} t_i^{\frac{1}{\gamma}} \right\} \quad (3.3)$$

où $t = (t_1, t_2, \dots, t_n)$ représentent les temps de survie, et où $\nu_i = 1$ si t_i correspond à la survenue d'un événement, et 0 si t_i correspond à une donnée censurée.

Afin de construire un modèle exponentiel par morceaux, il est nécessaire de construire un découpage fini du temps, $0 < s_1 < s_2 < \dots < s_J$, avec $s_J > t_i$ pour tout $i = 1, 2, \dots, n$. Le temps est donc divisé en J intervalles $(0, s_1], (s_1, s_2], \dots, (s_{J-1}, s_J]$. On suppose que le risque de base est constant dans le $j^{\text{ème}}$ intervalle, pour $t \in I_j = (s_{j-1}, s_j]$.

La fonction de vraisemblance des données historiques pour n patients dans le contexte d'un modèle exponentiel par morceaux est définie comme :

$$L(\lambda | D) = \prod_{i=1}^n \prod_{j=1}^J (\lambda_j)^{\delta_{ij} \nu_i} \exp \left\{ -\delta_{ij} \left[\lambda_j (t_i - s_{j-1}) + \sum_{g=1}^{j-1} \lambda_g (s_g - s_{g-1}) \right] \right\} \quad (3.4)$$

où λ_j avec $j = 1, \dots, J$ est le risque de base pour chaque intervalle, et $\delta_i = 1$ si le $i^{\text{ème}}$ sujet est en échec ou est censuré dans le $j^{\text{ème}}$ intervalle, et 0 sinon.

3.2.2 Mixture prior : Données agrégées sur l'effet du traitement

Le *mixture prior* est un mélange pondéré d'une composante informative et d'une composante vague. Cette approche a été présentée dans différents contextes notamment pour incorporer des données historiques dans un nouvel essai [90, 91]. La composante informative peut notamment être une distribution a priori prédictive dérivée d'une méta-analyse d'essais historiques [90].

3.2. Construction des distributions a priori

Un *mixture prior* à deux composantes, tel que présenté par Mutsvari et al. [91], nous permet de construire la distribution a priori de l'effet du traitement (β), défini comme le logarithme du HR. Il s'exprime par :

$$\pi(\beta|D_{ET}^H, \omega) = \omega \times \pi_H(\beta|D_{ET}^H) + (1 - \omega) \times \pi_v(\beta) \quad (3.5)$$

où $\pi_H(\beta|D_{ET}^H)$ est la composante informative résumant l'information historique (D_{ET}^H) sur le paramètre β .

Nous supposons que la distribution a priori du logarithme du HR est normale [47] :

$\pi_H(\beta|D_{ET}^H) \sim N(m_H, s_H^2)$ avec $m_H = \log(0.786)$ et $s_H^2 = 0.012$ (estimées à partir des données publiées de l'essai INT-0133, voir Section 3.1.4).

Une analyse de sensibilité a été réalisée en considérant des variances plus élevées, pour représenter des situations dans lesquelles la quantité d'information historique (nombre d'événements) est plus importante, comparable, ou plus faible que l'information attendue dans le nouvel essai. Nous avons défini ces valeurs comme étant $s_H^2 = s^2, 5s^2, et 15s^2$ équivalentes à 329, 66, et 22 événements, respectivement (nombre d'événements attendus calculé selon la formule de Schoenfeld [163]). Les principaux résultats présentés correspondent à $s_H^2 = s^2$, équivalent à 329 événements.

$\pi_v(\beta)$ est la composante vague qui assure une certaine robustesse face à un conflit entre les données actuelles et la distribution a priori en rendant les queues de distribution plus "lourdes".

Enfin, $\omega \in [0, 1]$ reflète la plausibilité a priori de la commensurabilité entre l'estimation de l'effet du traitement historique et l'effet du traitement du nouvel essai :

$\omega = 0$ indique une distribution a priori vague et $\omega = 1$ indique une distribution a priori informative basée sur les données historiques seulement. La Figure 3.2 illustre la manière dont la distribution du *mixture prior* basée sur les données historiques INT-0133 change avec le paramètre ω .

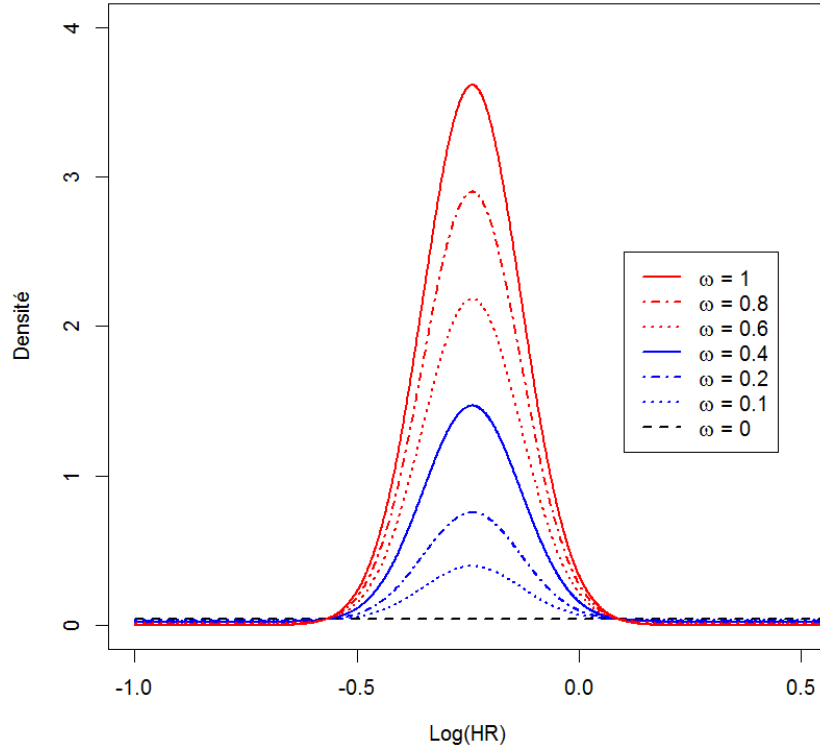


FIGURE 3.2 – **Distribution du *mixture prior* de l'effet du traitement pour différentes valeurs de ω**

Distribution du *mixture prior* où la composante informative est $\pi_H(\beta|D_{ET}^H) \sim N(\log(0.786), 0.012)$ et la composante vague est $\pi_v(\beta) \sim N(0, 10)$

3.2.3 Distribution a priori jointe

En supposant que les distributions a priori des paramètres du bras contrôle (θ) et de l'effet du traitement (β) sont indépendants, la distribution a priori jointe s'écrit comme le produit des distributions, c'est à dire :

$$\pi(\theta, \beta|D_C^H, D_{ET}^H, \alpha_0, \omega) \propto L(\theta|D_C^H)^{\alpha_0} \times \pi_0(\theta) \times [\omega \times \pi_H(\beta|D_{ET}^H) + (1 - \omega) \times \pi_v(\beta)] \quad (3.6)$$

3.3 Étude de simulations

L'objectif principal de l'étude de simulations est d'évaluer l'apport de l'approche Bayésienne combinant ou non des données historiques sur le bras contrôle et des données agrégées pour l'effet du traitement. Cette étude permet de mesurer les caractéristiques opératoires de l'analyse Bayésienne de la survie selon les poids (α_0, ω) alloués aux données historiques et ainsi de calibrer ces valeurs pour le nouvel essai. Il est ainsi possible de déterminer les valeurs pour α_0 et ω offrant le meilleur compromis entre augmentation de la puissance et de la précision, et augmentation du biais et de l'erreur de type I, selon différents scénarios représentant différents niveaux de commensurabilité entre les données historiques et les données du nouvel essai.

3.3.1 Définition des données individuelles historiques contrôles

Bien que les données historiques observées de l'essai OS2006 soient disponibles (données SARC-OS), nous avons décidé de travailler avec un jeu de données historiques simulé afin d'avoir un contrôle total à la fois sur le modèle de survie sous-jacent, mais également sur les paramètres de ce modèle. Cette première étude de simulations, incorporant des données historiques hypothétiques, a été complétée par une deuxième étude utilisant les données observées SARC-OS (cf. Figure 3.1) qui est détaillée Section 3.4.5.

Nous avons donc généré deux jeux de données historiques hypothétiques imitant les données de survie des patients SARC-OS mais provenant de deux distributions de survie différentes. Cela nous a permis d'évaluer les caractéristiques opératoires de la conception de l'essai Sarcome-13 dans :

- la situation idéale, où les données du bras contrôle du nouvel essai sont parfaitement commensurables avec les données historiques, en termes de forme de la distribution de survie.

- la situation où il y a des conflits de degré divers entre la distribution des données historiques, et celle des données contrôles du nouvel essai.

Ces deux jeux de données hypothétiques ont donc été générés de façon à être proches des données SARC-OS, tout en provenant d'un modèle de Weibull pour l'un, et d'un modèle exponentiel par morceaux pour l'autre.

En raison des variations aléatoires des simulations, en particulier dans le contexte d'échantillons de petite taille, nous avons suivi le processus suivant pour identifier ces deux jeux de données hypothétiques.

Dans un premier temps, pour chaque type de modèle, nous avons simulé 10 000 jeux de données en utilisant l'estimation du maximum de vraisemblance, obtenue en ajustant les données observées SARC-OS (en conservant la même taille d'échantillon, $N = 165$, voir Figure 3.1), afin de fixer les valeurs des paramètres du modèle de simulation.

Nous avons ensuite sélectionné de manière empirique une réplification ayant une fonction de survie estimée proche du modèle sous-jacent (similarité graphique), et des paramètres estimés proches de ceux des données historiques observées (voir Tableau 3.1).

Paramètres	Données SARC-OS	Jeux de données hypothétiques ^a	
		Weibull	Exponentiel par morceaux
Intercept (β_0)	1.683	1.682	-
Scale (γ)	1.100	1.098	-
Risques instantanés		-	
λ_1	0.224	-	0.223
λ_2	0.317	-	0.315
λ_3	0.082	-	0.082

TABLEAU 3.1 – Paramètres des modèles de Weibull et exponentiel par morceaux selon le jeu de données historiques utilisé

^a Jeux de données hypothétiques simulés selon un modèle de Weibull ou un modèle exponentiel par morceaux à partir des données SARC-OS

3.3. Étude de simulations

Les deux jeux de données hypothétiques retenus sont illustrés dans la Figure 3.3. Nous ferons référence par la suite à ces jeux de données en tant que HCW (données historiques contrôles générées à partir d'un modèle de Weibull) et HCP (données historiques contrôles générées à partir d'un modèle exponentiel par morceaux).

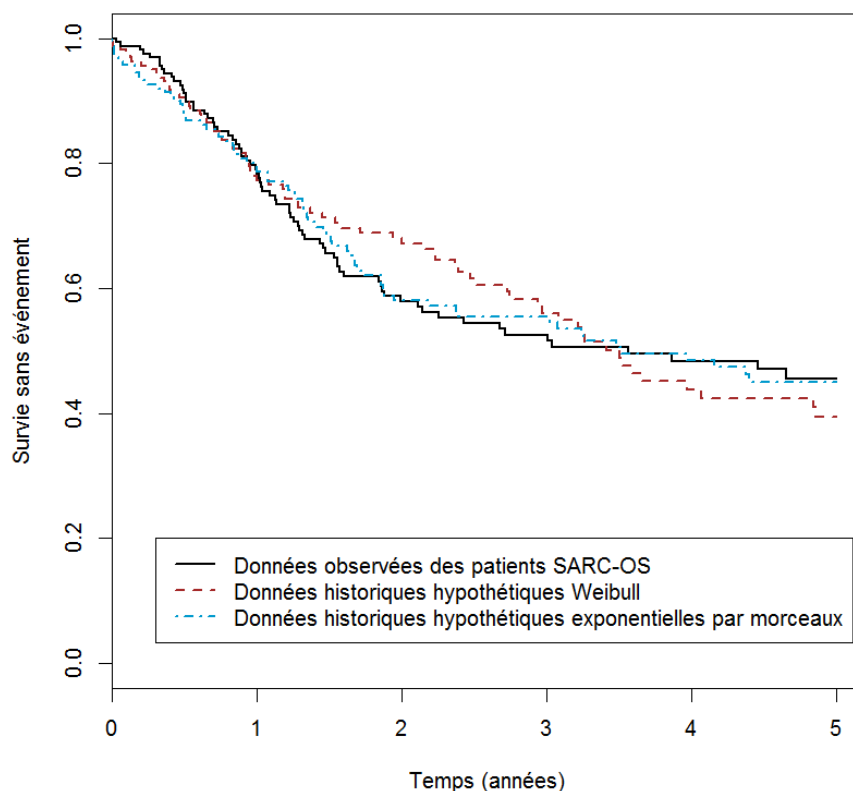


FIGURE 3.3 – Données historiques individuelles sur le bras contrôle considérées dans l'étude de simulations

Estimation de Kaplan-Meier des patients SARC-OS, correspondant au sous-groupe de patients de l'essai OS2006 satisfaisant les critères d'éligibilité de l'essai Sarcome-13 (courbe noire pleine), et estimation de Kaplan-Meier des deux jeux de données historiques hypothétiques retenus : l'un simulé à partir d'une distribution de Weibull (courbe marron en tirets) et l'autre simulé à partir d'une distribution exponentielle par morceaux (courbe bleue en tirets-pointillés).

3.3.2 Génération des données du nouvel essai

Les données du nouvel essai ont été générées afin de simuler des essais randomisés en deux groupes parallèles équilibrés, incluant 105 patients sur 3 ans avec 2 ans de suivi supplémentaire, avec des temps de survie individuels échantillonnés à partir d'une distribution de Weibull, ou d'une distribution exponentielle par morceaux.

Nous avons considéré différents niveaux de commensurabilité entre les données historiques contrôles et les données du nouvel essai, en termes de risque de base du groupe contrôle (pronostic plus ou moins péjoratif, représenté par les paramètres du modèle de survie du groupe contrôle), de taille d'effet du traitement évalué, et de forme du modèle de survie sous-jacent (Weibull *versus* exponentiel par morceaux).

Dans un premier temps (scénarios S1 à S24), nous avons supposé qu'il n'y avait pas de conflit de forme entre la distribution des données historiques du bras contrôle et celle des données du nouvel essai. Dans les scénarios S1 à S12 (respectivement, S13-S24), les données historiques HCW (respectivement, HCP) ont été utilisées, et les données du bras contrôle du nouvel essai ont été simulées à partir d'une distribution de Weibull (respectivement, exponentielle par morceaux). Trois cas ont été évalués :

- Scénarios S1 à S4 (respectivement, S13-S16) : nous avons fixé les paramètres du modèle de simulations, θ^W (respectivement, θ^P), comme étant égaux aux estimations obtenues en ajustant un modèle de Weibull (respectivement, exponentiel par morceaux) aux données SARC-OS (courbes vertes des Figures 3.4A et 3.4B).
- Scénarios S5 à S8 (respectivement, S17-S20) : les paramètres du modèle de simulations ont été fixés aux bornes inférieures (respectivement, supérieures) de l'intervalle de confiance à 95% pour θ^W (respectivement, θ^P) obtenues en ajustant les données SARC-OS (courbes rouges des Figures 3.4A et 3.4B) ; cette approche crée un "conflit négatif", où le pronostic des patients contrôles dans le nouvel essai est

moins bon que celui des patients contrôles historiques.

- Scénarios S9 à S12 (respectivement, S21-S24) : les paramètres du modèle de simulations ont été fixés aux bornes supérieures (respectivement, inférieures) de l'intervalle de confiance à 95% pour θ^W (respectivement, θ^P) (courbes bleues des Figures 3.4A et 3.4B), créant un "conflit positif", où le pronostic des patients contrôles dans le nouvel essai est meilleur que celui des patients contrôles historiques.

Les scénarios S25 à S32 représentent les cas où il existe un conflit entre les données historiques contrôles et les données du nouvel essai en matière de distribution de survie. Les scénarios S25 à S28 (respectivement, S29-S32) utilisent les données historiques contrôles générées à partir d'une distribution de Weibull (respectivement, exponentielle par morceaux), tandis que les nouvelles données sont générées à partir d'une distribution exponentielle par morceaux (respectivement, Weibull) (Figures 3.4C et 3.4D). Pour ces derniers scénarios les paramètres des fonctions de survie considérés correspondent aux paramètres issus de SARC-OS.

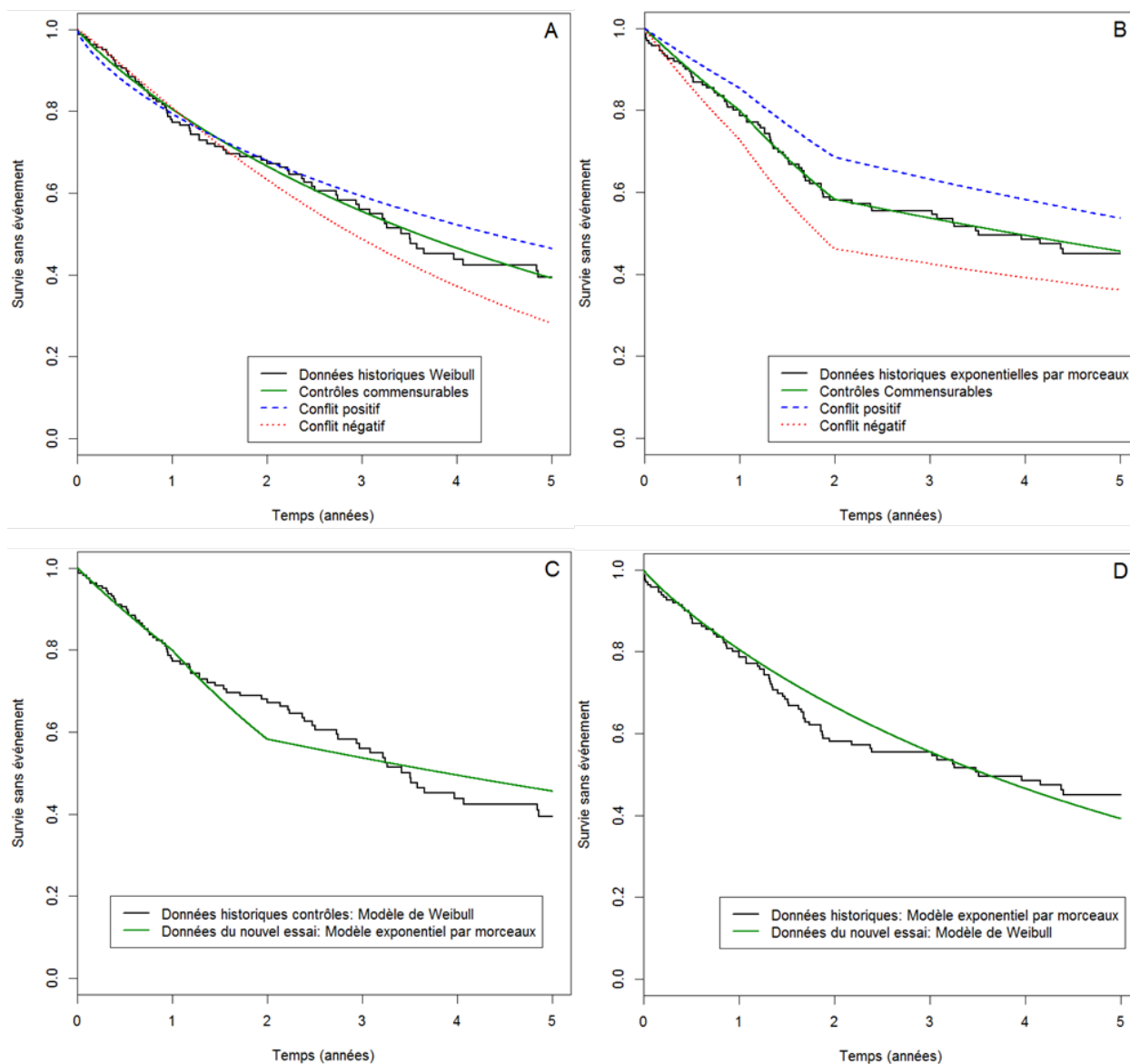


FIGURE 3.4 – **Distribution de la survie sans événement des données historiques et des données du bras contrôle du nouvel essai selon leur commensurabilité**

Sur chaque figure, la courbe noire représente les données contrôles historiques hypothétiques (estimations de Kaplan-Meier) simulées à partir d'une distribution de Weibull (HCW, Figures de gauche, A et C) et exponentielle par morceaux (HCP, Figures de droite, B et D). Les Figures A et B représentent la situation où il n'y a pas de conflit en termes de distribution de survie mais une possible non-commensurabilité en termes de paramètres (courbes vertes, rouges et bleues représentant des données commensurables, un conflit négatif, et un conflit positif, respectivement). Les Figures C et D représentent une non-commensurabilité en termes de distribution de survie (courbe verte).

3.3. Étude de simulations

Concernant l'effet du traitement du nouvel essai (β), quatre cas représentant différents degrés de commensurabilité avec les données historiques agrégées (D_{ET}^H) ont été considérés :

- Pas d'effet traitement : $\beta = \ln(1)$
Scénarios "effet nul" : S1, S5, S9, S13, S17, S21, S25, et S29
- Effet du traitement inférieur à l'effet du traitement historique dérivé de l'estimation globale de l'essai INT-0133 : $\beta = 0.5 \times \beta_H = 0.5 \times \ln(0.786) = \ln(0.886)$
Scénarios "effet décevant" : S2, S6, S10, S14, S18, S22, S26, et S30
- Effet du traitement égal à celui estimé par les données historiques :
 $\beta = \beta_H = \ln(0.786)$
Scénarios "effet historique" : S3, S7, S11, S15, S19, S23, S27, et S31
- Effet du traitement égal à l'effet cible du calcul de puissance de l'essai Sarcome-13, effet résolument optimiste, qui est supérieur à l'effet du traitement historique :
 $\beta = \ln(0.55)$
Scénarios "effet espéré" : S4, S8, S12, S16, S20, S24, S28, et S32

Nous avons donc considéré huit configurations représentant différents types de conflits entre la distribution de survie de l'essai futur et des données contrôles historiques. De plus, nous avons évalué quatre situations pour l'effet du traitement dans le nouvel essai, plus ou moins différent de l'effet des données historiques.

Au total, nous avons donc établi 32 scénarios de simulations reflétant différents degrés de commensurabilité entre les informations historiques, et les données du nouvel essai (Tableau 3.2, voir Appendix E pour la représentation graphique des différents scénarios).

Pour chaque scénario, 5 000 essais ont été simulés avec un échantillon fixe de 105 patients, comme cela a été planifié dans l'essai Sarcome-13, avec une randomisation équilibrée (1:1) entre les bras de l'essai, et un taux de censure uniforme de 5%.

Chapitre 3. Proposition d'un design Bayésien avec incorporation de données historiques individuelles et agrégées pour un essai clinique

Scénarios	Distribution contrôle historique ^a	Génération des nouvelles données		
		Distribution de survie ^b	Paramètres	
			Bras contrôle	Effet traitement ^c
S1				Nul
S2			Contrôles	Décevant
S3			commensurables	Historique
S4				Espéré
S5				Nul
S6	Weibull (HCW)	Weibull	Conflit	Décevant
S7			négatif	Historique
S8				Espéré
S9				Nul
S10			Conflit	Décevant
S11			positif	Historique
S12				Espéré
S13				Nul
S14			Contrôles	Décevant
S15			commensurables	Historique
S16				Espéré
S17				Nul
S18	Exponentielle	Exponentielle	Conflit	Décevant
S19	par morceaux (HCP)	par morceaux	négatif	Historique
S20				Espéré
S21				Nul
S22			Conflit	Décevant
S23			positif	Historique
S24				Espéré
S25				Nul
S26	Weibull (HCW)	Exponentielle	Contrôles	Décevant
S27		par morceaux	commensurables	Historique
S28				Espéré
S29				Nul
S30	Exponentielle	Weibull	Contrôles	Décevant
S31	par morceaux (HCP)		commensurables	Historique
S32				Espéré

TABLEAU 3.2 – Résumé des 32 scénarios considérés pour l'étude de simulations

^a Distribution de survie utilisée pour générer les données historiques contrôles

^b Distribution de survie utilisée pour générer les données du nouvel essai

^c Les effets nul, décevant, historique et espéré correspondent à un hazard ratio de 1, 0.886, 0.786, et 0.55 dans le nouvel essai, respectivement

HCW : Données historiques contrôles générées à partir d'un modèle de Weibull

HCP : Données historiques contrôles générées à partir d'un modèle exponentiel par morceaux

3.3.3 Stratégie d'analyse des données simulées

Les données générées selon les scénarios S1-S12 et S13-S24, pour lesquelles il n'y avait pas de conflit en termes de forme de distribution de survie, ont été analysées en ajustant un modèle Bayésien de Weibull ou exponentiel par morceaux, respectivement.

Concernant les scénarios S25-S32, dans lesquels les données historiques contrôles et les données du nouvel essai sont tirées de distributions différentes, nous avons évalué l'impact du modèle d'analyse en comparant les résultats obtenus lorsque les données étaient analysées au travers d'un modèle cohérent, d'une part avec la distribution sous-jacente des données historiques contrôles, et d'autre part avec la distribution utilisée pour générer les données du nouvel essai.

Lors de l'analyse de l'essai, nous avons défini les distributions a priori, pour chaque modèle considéré, comme suit :

— Modèle de Weibull :

$$\beta_0 \sim N(0, 10\,000)$$

$$\gamma \sim \text{IGamma}(0.0001, 0.0001)$$

— Modèle exponentiel par morceaux :

$$\lambda_1, \lambda_2, \text{ et } \lambda_3 \text{ (échelle logarithmique)} \sim N(0, 10\,000)$$

Pour évaluer l'impact de l'incorporation des données historiques nous avons réalisé les analyses sous différentes configurations concernant les poids du *power prior* et du *mixture prior*, α_0 et ω :

— $\alpha_0 \in \{0, 0.3, 0.6, 1\}$

— $\omega \in \{0, 0.1, 0.2, 0.4, 0.6, 0.8, 1\}$

Un pas plus étroit, en particulier sur les petites valeurs, a été choisi pour le paramètre ω pour évaluer plus finement l'impact de la prise en compte des données historiques portant sur l'effet du traitement.

Afin de définir la variance de la composante vague du *mixture prior*, nous avons réalisé une première étude de simulations. Plusieurs variances ont été envisagées : 1 000, 100 et 10. Nos simulations ont cependant mis en évidence qu'une variance trop importante ne permettait pas de pondérer correctement les données (voir Figure 3.5). En effet, dans ce cas, la distribution a priori aura des queues de distribution très lourdes, plaçant la masse de la distribution sur des valeurs de l'effet du traitement très grandes et improbables. C'est la raison pour laquelle nous avons choisi pour nos simulations que : $\pi_v(\beta) \sim N(0, 10)$.

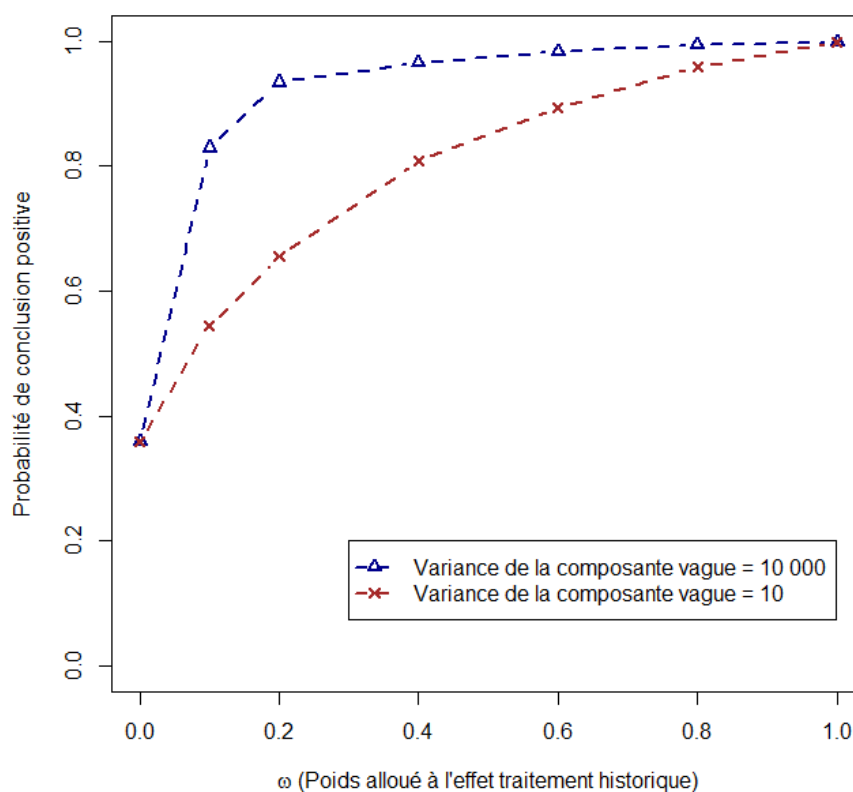


FIGURE 3.5 – Effet de la variance de la composante vague du *mixture prior* sur la pondération des données historiques sur l'effet du traitement dans le scénario S3

Effet de la variance de la composante vague du *mixture prior* sur la pondération des données historiques sur l'effet du traitement (exemple de la puissance dans le scénario S3) selon que celle-ci est égale à 10 ou 10 000.

Les lois a posteriori des paramètres d'intérêt ont été estimées en utilisant un processus de type *Markov Chain Monte Carlo*. Nous avons réalisé une chaîne, en utilisant l'algorithme de Metropolis-Hastings [52] pour 20 000 itérations avec un *burn-in* de 5 000 itérations, laissant 15 000 itérations pour les inférences a posteriori. La convergence de la distribution a été évaluée par le test diagnostique de Geweke [53]. Au terme de chaque essai analysé selon cette approche, la distribution a posteriori a été résumée par la moyenne et l'écart-type du Log-HR estimé.

Nous avons évalué les caractéristiques opératoires fréquentistes du design Bayésien proposé en supposant que la règle suivante sera utilisée pour prendre la décision finale : L'utilisation conjointe du mifamurtide et de la chimiothérapie post-opératoire sera jugée supérieure à la chimiothérapie post-opératoire seule, si la probabilité a posteriori d'un HR inférieur à 1 est supérieure ou égale à 0.9 ($P(HR < 1) \geq 0.9$).

3.3.4 Critères d'évaluation de l'impact des données historiques sur l'effet du traitement estimé

Pour chaque scénario, nous avons estimé le biais de l'estimation de l'effet du traitement a posteriori comme étant la différence entre la moyenne des 5 000 estimations de β a posteriori (moyenne de la loi a posteriori) et le véritable effet du traitement défini dans le scénario.

L'écart-type empirique (SD) a été calculé comme une mesure de la dispersion et de la précision de l'estimation de β (effet traitement), et la racine carrée de l'erreur quadratique moyenne (RMSE) comme mesure d'exactitude globale (compromis entre le biais et la précision de l'estimation). Ils sont définis comme :

$$SD = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (\hat{\beta}_i - \bar{\beta})^2}$$

$$RMSE = \sqrt{E((\hat{\beta} - \beta)^2)}$$

En comptant le nombre de conclusions positives parmi les essais simulés, nous avons pu calculer l'erreur fréquentiste de type I de formulation unilatérale (dans les scénarios à effet nul, où $HR = 1$), et la puissance fréquentiste d'une décision Bayésienne (dans tous les autres scénarios, dans lesquels le nouveau traitement est supérieur au contrôle : $HR < 1$).

Toutes les simulations et les analyses Bayésiennes ont été réalisées en utilisant un programme écrit sous R 3.3.1 [164], en appelant les packages LearnBayes [165], et MCMCpack [166].

3.4 Résultats

Dans un premier temps, les résultats des simulations utilisant les jeux de données historiques hypothétiques pour le bras contrôle sont présentés.

Dans un deuxième temps, les simulations basées sur les données historiques contrôles réelles (données SARC-OS) sont présentées.

3.4.1 Impact de l'inclusion de données historiques agrégées portant sur l'effet du traitement seulement

L'impact de l'inclusion de données historiques agrégées sur l'effet du traitement uniquement (excluant les données historiques sur le bras contrôle) peut être évalué en observant les résultats des simulations lorsque $\alpha_0 = 0$. Les Figures 3.6 à 3.8 résument les principaux résultats des simulations pour les scénarios S1 à S12 quant aux erreurs de type I et II, au biais, et à la RMSE de l'estimation de l'effet du traitement a posteriori.

Dans ces scénarios, les données historiques contrôles ont été générées à partir d'une distribution de Weibull. Les conflits entre les données historiques et les données du

nouvel essai proviennent soit de différences entre les paramètres correspondants des distributions de Weibull, soit de différences entre l'effet réel du traitement du nouvel essai et l'effet du traitement historique.

Les résultats détaillés des scénarios S1 à S12 sont disponibles dans les Tableaux F1 à F3 de l'Annexe F. Des résultats similaires ont été observés lorsque les données étaient générées à partir d'une distribution exponentielle par morceaux (scénarios S13 à S24). Ces résultats sont disponibles dans les Tableaux F4 à F6 de l'Annexe F.

L'effet de l'augmentation de ω est illustré par la courbe noire sur les Figures 3.6 à 3.8, où $\alpha_0 = 0$. Dans les scénarios S1-S4, S5-S8 et S9-S12, les données du bras contrôle du nouvel essai ont été générées à partir de modèles de Weibull ayant différents risques de base. Ces différences n'ont pas d'impact majeur. De ce fait, des résultats similaires sont observés parmi les scénarios "Effet nul" (S1, S5, et S9). La même observation s'applique aux scénarios représentant un "Effet décevant" (S2, S6, et S10), et un "Effet espéré" (S4, S8, et S12).

Incorporer des données historiques sur l'effet du traitement conduit à des gains de puissance importants, d'une part lorsque l'effet du traitement dans le nouvel essai est inférieur à l'estimation historique (la puissance augmente de 20.8% pour $\omega = 0$ à 98.9% pour $\omega = 1$ dans le scénario S2), et d'autre part lorsqu'il est plus important que l'estimation historique (la puissance augmente de 79.4% pour $\omega = 0$ à 100% pour $\omega = 1$ dans le scénario S4). Dans la plupart des scénarios, les gains de puissance les plus importants sont obtenus en augmentant ω de 0 à 0.2 ou 0.4 ; augmenter ω au-delà entraîne un gain moins important. Les augmentations de puissance observées à travers les différents scénarios doivent être contre-balançées avec le risque d'augmenter l'erreur de type I si, en réalité, le nouveau traitement n'est pas meilleur que le traitement standard dans le nouvel essai : dans le scénario "Effet nul" S1, l'erreur de type I augmente rapidement avec ω , pour atteindre 96.9% pour $\omega = 1$.

Dans le scénario où le nouvel effet du traitement est égal à l'estimation historique (scénario S3), l'incorporation de données historiques sur l'effet du traitement n'entraîne pas de biais. Si l'effet du traitement dans le nouvel essai est moins bon que celui estimé dans les données historiques (scénarios S1-S2), le nouvel effet est alors surestimé (biais négatif sur $\beta = \text{Log}(HR)$) et la magnitude du biais augmente avec ω . À l'inverse, si l'effet du traitement dans le nouvel essai est supérieur à celui estimé dans les données historiques (scénario S4), l'incorporation de données historiques entraîne une sous-estimation (biais positif sur β) du nouvel effet du traitement.

Lorsque $\omega = 0$, l'écart-type empirique moyen de l'estimation a posteriori de l'effet du traitement est légèrement différent selon les scénarios en raison des différents nombres d'événements attendus dans le nouvel essai (moins d'événements attendus lorsqu'il y a un effet du traitement plus important, ou lorsque le risque de base de d'événements est plus petit).

Malgré cela, dans tous les scénarios, la dispersion moyenne de l'estimation a posteriori diminue avec ω . En revanche, la manière dont la RMSE change avec ω dépend du scénario. Par exemple, dans les scénarios S2 et S3 (effet décevant et historique), la RMSE diminue rapidement lorsque ω augmente : la variation relative de la RMSE est de -59% dans le scénario S2 (due au compromis entre augmentation du biais et diminution de la dispersion de l'estimation) et de -86% dans le scénario S3 lorsque ω augmente de 0 à 1. Dans le scénario "Effet nul" (S1), une diminution plus faible de la RMSE est observée (variation relative de -21% lorsque ω augmente de 0 à 1). Dans le scénario "Effet espéré" (S4), de légères variations de la RMSE sont observées lorsque ω varie entre 0 et 1, mais, tandis qu'une faible diminution est observée lorsque ω augmente de 0 à 0.2, une petite augmentation est observée lorsque ω est supérieur à 0.4, reflétant le compromis entre une diminution de la dispersion de l'estimation, et une augmentation du biais.

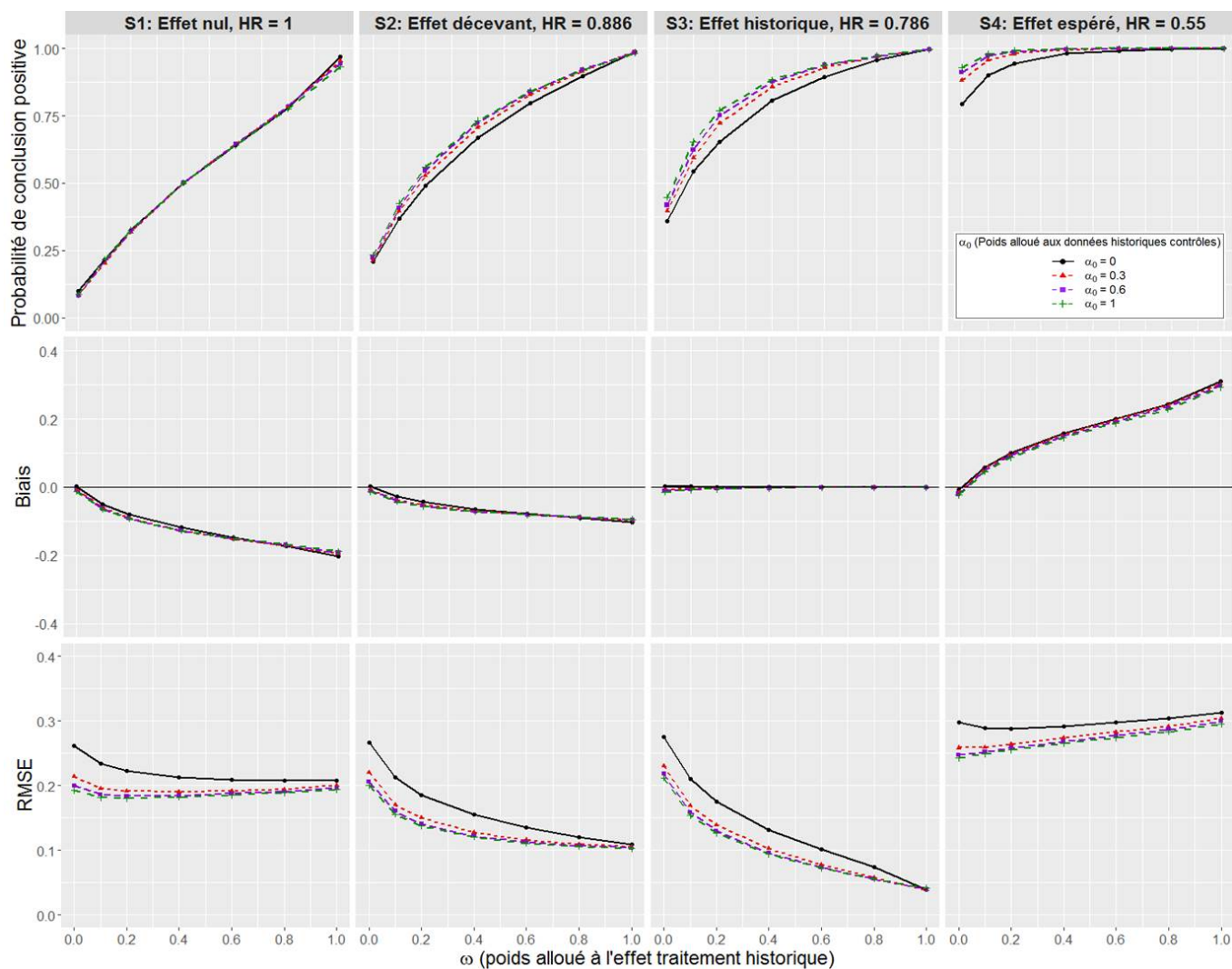


FIGURE 3.6 – Impact de α_0 et ω sur les caractéristiques opératoires pour les scénarios S1-S4

Une distribution de Weibull est utilisée pour générer les données historiques contrôles et les données du nouvel essai, avec s_H^2 équivalent à 329 événements. Dans ces scénarios (S1-S4), les données contrôles historiques et les données contrôles du nouvel essai sont commensurables.

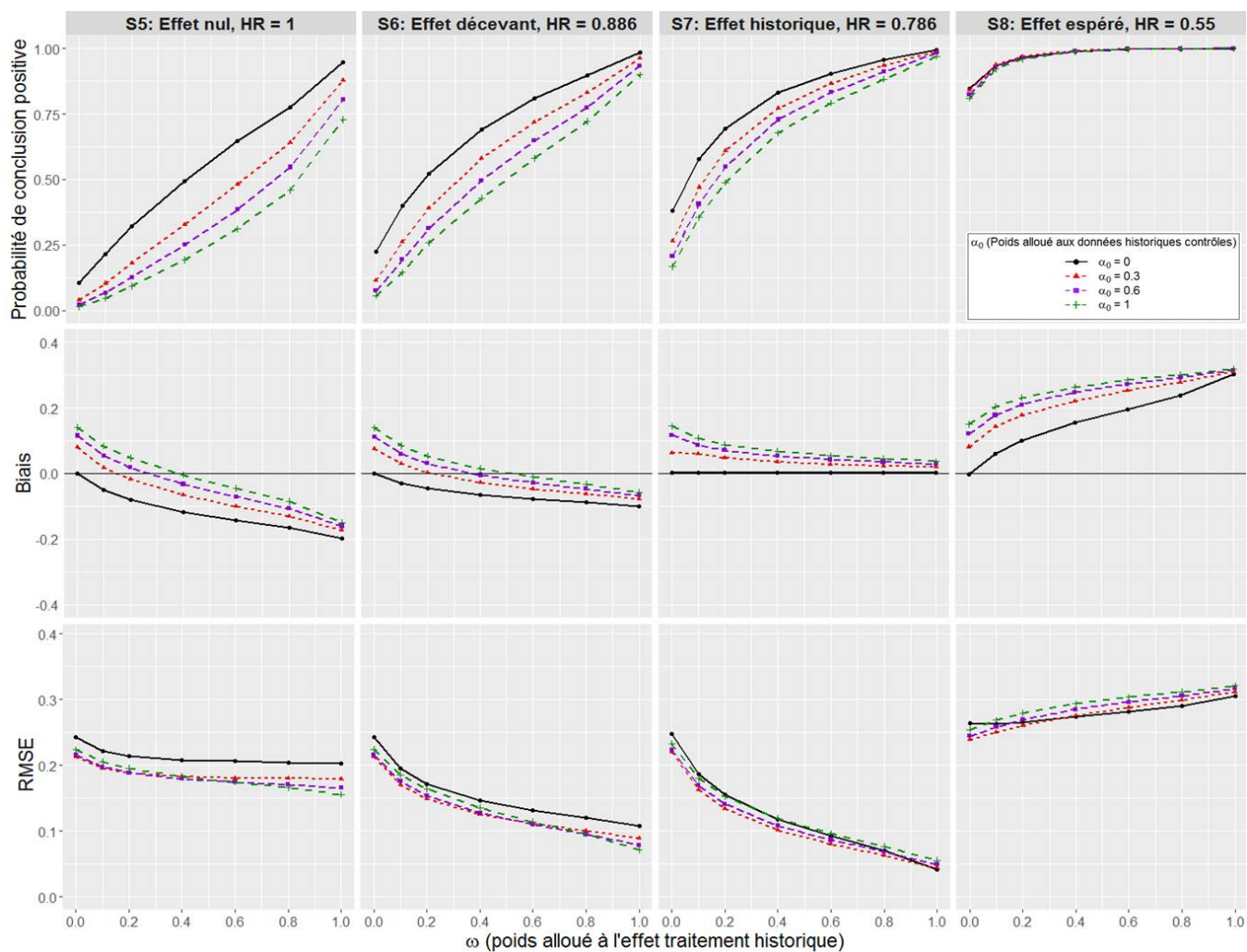


FIGURE 3.7 – Impact de α_0 et ω sur les caractéristiques opératoires pour les scénarios S5-S8

Une distribution de Weibull est utilisée pour générer les données historiques contrôles et les données du nouvel essai, avec s_H^2 équivalent à 329 événements. Dans ces scénarios (S5-S8), un conflit négatif (où les patients du groupe contrôle du nouvel essai ont un moins bon pronostic que les patients du groupe contrôle historique) existe .

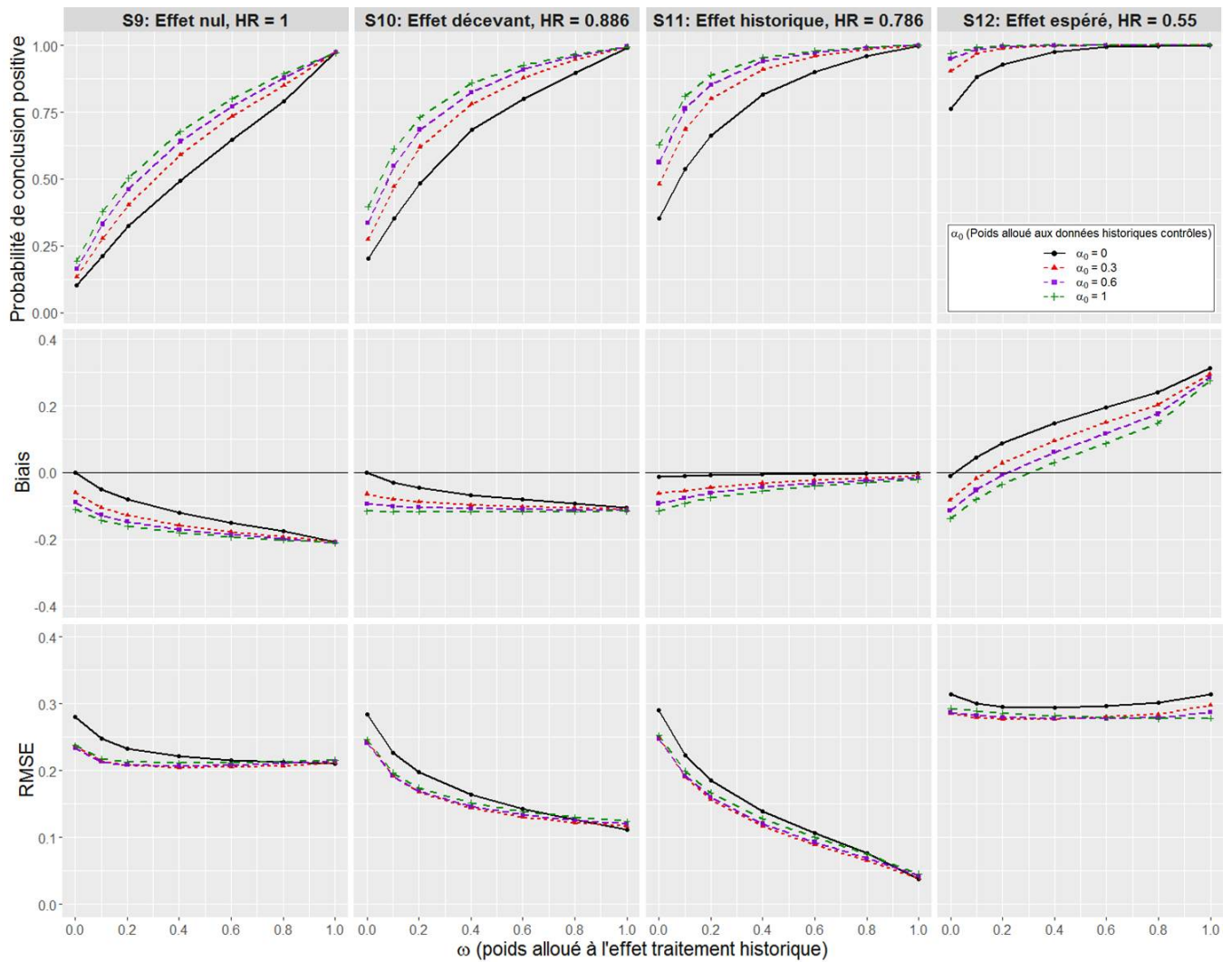


FIGURE 3.8 – Impact de α_0 et ω sur les caractéristiques opératoires pour les scénarios S9-S12

Une distribution de Weibull est utilisée pour générer les données historiques contrôles et les données du nouvel essai, avec s_H^2 équivalent à 329 événements. Dans ces scénarios (S9-S12), un conflit positif (où les patients du groupe contrôle du nouvel essai ont un meilleur pronostic que les patients du groupe contrôle historique) existe.

Choix de la valeur de ω pour l'essai Sarcome-13

Le choix du poids à attribuer aux données historiques sur l'effet du traitement doit être fait en considérant l'impact de l'augmentation de ω sur le biais, la dispersion (SD), et l'exactitude globale (RMSE) de l'estimateur de l'effet du traitement, ainsi que sur la puissance et l'erreur de type I de l'essai.

Par conséquent, nous recommandons de fixer $\omega = 0.1$ lorsque $\alpha_0 = 0$, si la variance de la composante informative du *mixture prior* pour β est équivalente à 329 événements. Cette valeur de ω conduit à une erreur de type I de 21.6% (S1), tout en permettant un gain substantiel de puissance lorsque ω augmente de 0 à 0.1 (de 35.9% à 54.5% dans le scénario S3, par exemple), et en exactitude de l'estimation (la RMSE diminue de 0.275 à 0.210 dans le scénario S3).

Analyse de sensibilité en cas de données historiques moins informatives

Différentes recommandations pour ω peuvent s'appliquer en fonction de la quantité d'informations historiques disponible pour l'effet du traitement.

Lorsque $s_H^2 = s^2$, $5s^2$, ou $15s^2$ (équivalent à 329, 66 et 22 événements respectivement), nous observons une tendance similaire dans les propriétés de l'estimateur de l'effet du traitement, et des caractéristiques opératoires de l'essai, lorsque ω augmente de 0 à 1. Néanmoins, l'impact des changements de ω diminue lorsque s_H^2 augmente. Cela est particulièrement vrai pour la puissance, comme l'illustre la Figure 3.9 pour les scénarios S1 à S4. Par exemple, pour le scénario S3, lorsque $s_H^2 = s^2$, la puissance augmente de 35.9% à 99.7% en augmentant ω de 0 à 1, alors qu'elle augmente à 54.2% lorsque $s_H^2 = 5s^2$, et à 40.7% lorsque $s_H^2 = 15s^2$ (les détails des résultats pour tous les scénarios sont disponibles dans les Tableaux G1 à G6 de l'Annexe G).

Pour toutes valeurs de s_H^2 , le choix optimal pour ω doit équilibrer les objectifs concurrents d'augmentation de la puissance, et de contrôle du biais.

Par exemple, si l'on suppose que l'information historique sur β est telle que $s_H^2 = 5s^2$, alors dans ce cas, si la distribution de survie est telle que spécifiée dans les scénarios S2-S4, pour obtenir le même gain de puissance que celui atteint en choisissant $\omega = 0.1$ lorsque $s_H^2 = s^2$, nous devons fixer $\omega = 1$.

Cependant, les risques auxquels on s'expose en fixant $\omega = 1$ lorsque $s_H^2 = 5s^2$ sont plus élevés que ceux associés à $\omega = 0.1$ lorsque $s_H^2 = s^2$. En effet, nous observons un biais plus important dans l'estimateur de l'effet du traitement lorsque les données historiques et les données du nouvel essai ne sont pas commensurables.

Par exemple, dans le scénario S2, le biais est estimé à -0.0275 lorsque $s_H^2 = s^2$ et $\omega = 0.1$ contre -0.0627 lorsque $s_H^2 = 5s^2$ et $\omega = 1$; dans le scénario S4, le biais est de 0.0582 contre 0.2014. Quand la quantité d'information historique est réduite ($s_H^2 = 15s^2$), les gains de puissance possibles en incorporant les informations historiques sur β sont négligeables, bien que le biais augmente toujours avec ω .

Cela contraste avec le fait que des gains en exactitude globale sont toujours possibles en augmentant ω , même pour de grandes valeurs de la variance s_H^2 . On observe en effet, pour le scénario S3 par exemple, une variation relative de -86% lorsque $s_H^2 = s^2$ en augmentant ω de 0 à 1, de -55% lorsque $s_H^2 = 5s^2$, et de -29% lorsque $s_H^2 = 15s^2$.

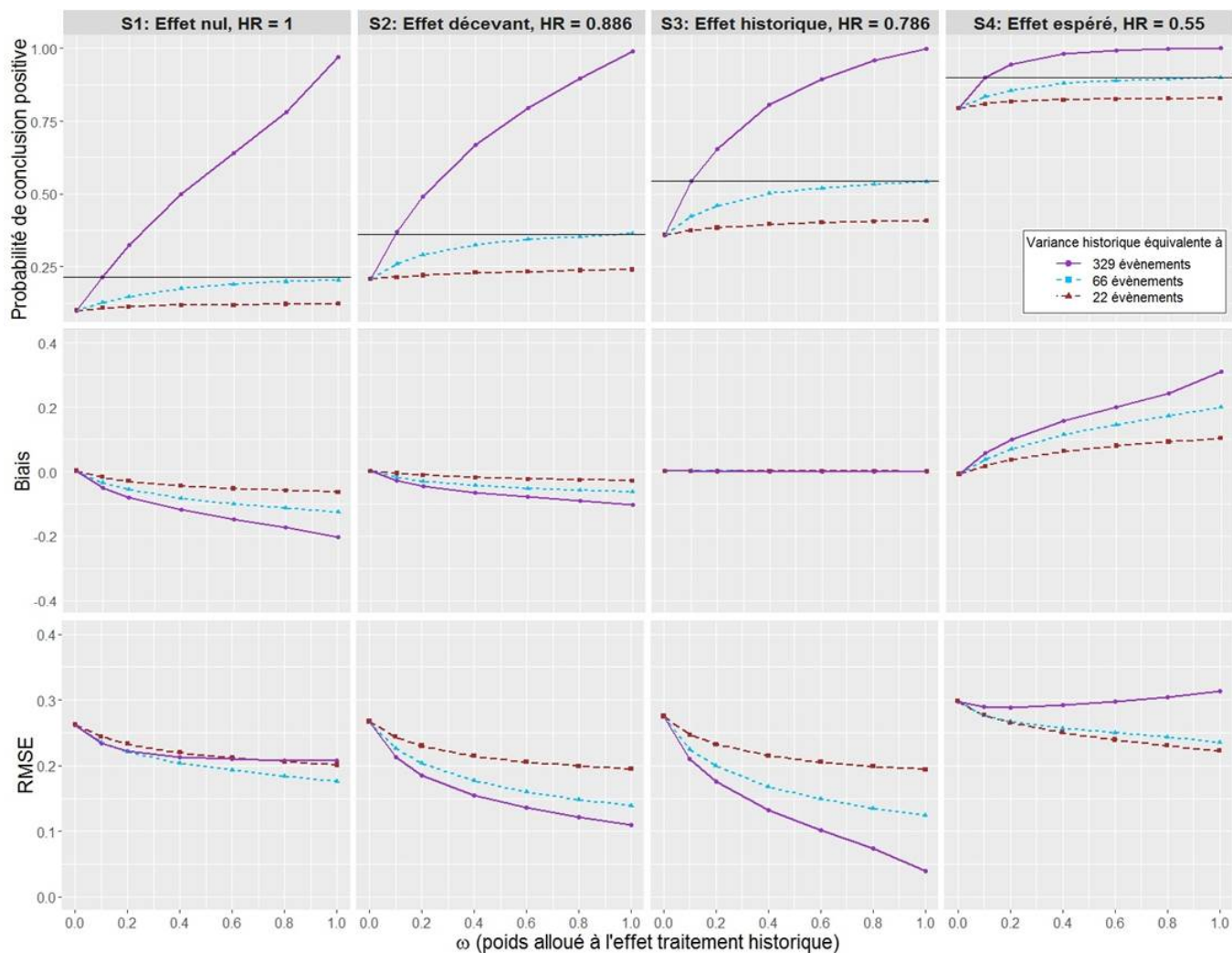


FIGURE 3.9 – Impact de ω , pour $\alpha = 0$ et pour différentes variances de la composante informative du *mixture prior*, sur les caractéristiques opératoires pour les scénarios S1-S4

Une distribution de Weibull est utilisée pour les données historiques et les données du nouvel essai, avec différentes valeurs pour s_H^2 équivalentes à 329, 66, et 22 événements. Nous supposons que les données contrôles historiques et celles du nouvel essai sont commensurables. La ligne horizontale représente la valeur du critère d'évaluation pour $\alpha_0 = 0$ et $\omega = 0.1$.

3.4.2 Impact de l'inclusion de données historiques individuelles portant sur le bras contrôle seulement

Cette section résume tout d'abord les résultats pour les scénarios S1 à S12 (qui supposent que les données historiques et celles du nouvel essai sont générées à partir d'un modèle de Weibull), pour $\omega = 0$, et différentes valeurs de α_0 . Ces résultats peuvent être lus en observant les ordonnées à l'origine des courbes présentées dans les Figures 3.6 à 3.8 (les détails de ces résultats sont disponibles dans les Tableaux F1 à F3 de l'Annexe F).

Données contrôles commensurables

Lorsque seules les données historiques contrôles sont intégrées dans l'analyse du nouvel essai, cela conduit à une augmentation limitée de la puissance lorsque les données historiques et celles du nouvel essai sont commensurables (scénarios S1-S4, Figure 3.6). L'importance du gain de puissance possible en augmentant α_0 de 0 à 1 dépend de l'effet réel du traitement, avec une augmentation de 20.8% à 23.3% dans le scénario S2 ; de 35.9% à 44.8% dans le scénario S3 ; et de 79.4% à 92.9% dans le scénario S4. La puissance est augmentée en variant α_0 de 0 à 0.3, avec des gains possibles, mais limités, pour des valeurs de α_0 plus élevées.

L'augmentation de α_0 est aussi associée à une légère amélioration du contrôle de l'erreur de type I ; par exemple, le taux de faux positifs diminue de 10% à 8.7%, lorsque α_0 augmente de 0 à 1, dans le scénario S1. Les principaux avantages de l'incorporation de données historiques contrôles sont une diminution de la dispersion, et un gain en termes d'exactitude globale de l'estimation de l'effet du traitement (la RMSE diminue de 0.275 à 0.212 dans le scénario S3).

De manière inattendue, le biais de l'estimation augmente légèrement avec α_0 . Cela est probablement expliqué par la taille limitée de l'échantillon du nouvel essai. Pour vérifier cette hypothèse, nous avons étendu nos simulations afin de considérer le cas où

le nouvel essai recrutait 2 000 patients (données non présentées). Dans ces scénarios supplémentaires, nous n'avons observé aucune augmentation du biais avec α_0 , ce qui confirme que le faible biais observé ici, lorsque le nouvel essai ne recrute que 105 patients, est dû à la petite taille de l'échantillon.

Conflit négatif

Concernant à présent les scénarios où la survie des patients du groupe contrôle du nouvel essai est moins bonne que la survie des patients du groupe contrôle historique (scénarios S5 à S8, Figure 3.7), nous pouvons observer une réduction de la probabilité de conclure à un essai positif dans tous les scénarios, conduisant à un meilleur contrôle de l'erreur de type I (passant de 10.4% à 1.4% quand $\alpha_0 = 0$ et 1, respectivement dans le scénario S5), mais aussi à une perte de puissance dans les autres scénarios. En attribuant plus de poids aux données historiques, on note que la puissance diminue de 22.4% à 5.6% (variation relative : -75%), de 38.2% à 16.8% (-56%), et de 84.9% à 81.1% (-4.5%), dans les scénarios "Effet décevant" (S6), "Effet historique" (S7), et "Effet espéré" (S8), respectivement. La perte de puissance observée est due à la sous-estimation de l'effet du traitement, l'importance du biais augmentant avec α_0 . En effet, l'augmentation de α_0 revient à augmenter les données du groupe contrôle du nouvel essai avec des pseudo-observations dont le pronostic est meilleur. Cela réduit donc la différence entre les bras expérimental et contrôle dans l'analyse du nouvel essai.

La diminution de la dispersion de l'estimation, observée parmi les scénarios S5-S8, est similaire à celle observée dans les scénarios S1-S4, lorsque les données contrôles sont commensurables. L'augmentation du biais et la diminution de la dispersion de l'estimation observées lorsque α_0 augmente, entraînent une légère amélioration de la RMSE pour $\alpha_0 = 0.3$ mais qui disparaît pour de plus grandes valeurs de α_0 . Par conséquent, dans ce contexte, il y a peu d'avantages à intégrer des données historiques contrôles en matière de RMSE et nous observons un effet délétère en matière de puissance.

Conflit positif

Concernant à présent les scénarios où la survie des patients du groupe contrôle du nouvel essai est meilleure que la survie des patients du groupe contrôle historique (scénarios S9 à S12, Figure 3.8), nous observons des résultats similaires en matière de dispersion de l'estimation de l'effet du traitement.

Dans ces scénarios, l'effet du traitement est sur-estimé. En effet, augmenter le nombre d'observations du groupe contrôle du nouvel essai par des données historiques ayant un moins bon pronostic, augmente l'estimation de la différence entre les traitements. Cela conduit à une erreur de type I qui augmente avec α_0 de 10.2% à 19.4% (Figure 3.8), et à une puissance qui augmente également.

Résultats pour une distribution exponentielle par morceaux

Dans l'ensemble, des résultats similaires sont observés lorsque les données historiques et les données du nouvel essai suivent une distribution exponentielle par morceaux (scénarios S13 à S24). Les résultats détaillés sont présentés dans les Tableaux F4 à F6 dans l'Annexe F.

Lorsqu'un petit échantillon est considéré, le biais dans l'estimation de l'effet du traitement observé lorsque les données contrôles historiques et celles du nouvel essai sont commensurables est plus marqué lorsque les résultats proviennent d'un modèle exponentiel par morceaux comparativement à un modèle de Weibull. Cependant, ce biais disparaît lorsque l'on considère des tailles d'échantillon plus importantes dans le nouvel essai (voir Tableau 3.3).

α_0	Taille d'échantillon	
	N = 105	N = 2 000
0	0.0029	0.0005
0.3	-0.0251	0.0003
0.6	-0.0330	0.0006
1	-0.0381	0.0001

TABLEAU 3.3 – Modèle exponentiel par morceaux : impact de α_0 en termes de biais pour différentes tailles d'échantillon (scénario S3)

3.4.3 Incorporation de données historiques individuelles contrôles et de données historiques agrégées sur l'effet du traitement

Dans un premier temps, les résultats présentés dans cette section suppose que les données historiques contrôles et celles du nouvel essai suivent une distribution de Weibull, et que la variance de la composante informative du *mixture prior* pour l'effet du traitement est équivalente à l'observation de 329 événements ($s^2 = 0.012$).

Dans ce contexte, nous supposons que nous incorporons à la fois les données historiques contrôles, et les données historiques sur l'effet du traitement dans l'analyse du nouvel essai en faisant varier α_0 et ω ($\alpha_0 > 0$ et $\omega > 0$).

Nous constatons que les changements dans la probabilité de conclure à un essai positif, l'écart-type empirique, et la RMSE de l'estimation a posteriori de l'effet du traitement semblent être largement déterminés par les variations de ω plutôt que par celles de α_0 . Cela se traduit par des courbes très similaires lorsque $\alpha_0 > 0$, surtout pour les scénarios S1-S4 (Figures 3.6 à 3.8).

La manière dont les caractéristiques opératoires changent lorsque ω augmente est stable parmi les différentes valeurs de α_0 . Dans de nombreux scénarios, l'impact de α_0 sur les performances des critères d'évaluation est moins important lorsque ω est grand.

Cela s'explique par le fait que ω est directement relié à l'information sur l'effet du traitement. Dans le scénario S3, lorsque toutes les données historiques (groupe contrôle et effet du traitement) sont commensurables avec les données du nouvel essai, la puissance est respectivement de 35.9% quand $\alpha_0 = 0$ et $\omega = 0$; 44.8% quand $\alpha_0 = 1$ et $\omega = 0$; et 99.7% quand $\omega = 1$, indépendamment de la valeur de α_0 .

Par ailleurs, nous observons une inflation globale de l'erreur de type I avec l'augmentation de ω , quel que soit le niveau de commensurabilité entre les données contrôles historiques et celles du nouvel essai, et quelles que soient les valeurs de α_0 (scénarios S1, S5, S9 où un "Effet nul" est observé ; Figures 3.6 à 3.8).

Nous remarquons toutefois un impact positif de l'augmentation de α_0 dans le cas d'un conflit négatif (scénario S5) : par exemple, pour $\omega = 1$, l'erreur de type I varie de 95% pour $\alpha_0 = 0$, à 73% pour $\alpha_0 = 1$. Cependant, on observe une inflation rapide de l'erreur de type I lorsque l'on augmente ω contrebalançant le faible gain obtenu avec l'augmentation de α_0 .

Choix de la valeur des paramètres de pondération pour l'essai Sarcome-13

En utilisant la configuration $\alpha_0 = 0$ et $\omega = 0$ comme référence, nous pouvons conclure qu'incorporer des données historiques en fixant $\alpha_0 = 0.3$ et $\omega = 0.1$ est un compromis acceptable.

Cela augmente la puissance de 20.8% à 39.5% dans le scénario "effet décevant" (S2) ; de 35.9% à 61.3% dans le scénario "Effet historique" (S3) ; et de 79.4% à 95.7% dans le scénario "Effet espéré" (S4).

La configuration $\alpha_0 = 0.3$ et $\omega = 0.1$ conduit également à une erreur de type I de 20.4% (S1) lorsque les données contrôles sont commensurables, mais qui est jugée acceptable dans le cadre de l'essai Sarcome-13. En effet, le scénario d'absence d'effet du mifamurtide sur

l'EFS est jugé peu probable par les investigateurs. De plus, dans leurs travaux, Bayar et al. [35] concluent que la réalisation consécutive de plusieurs essais randomisés de petite taille, acceptant une erreur de type I relâchée, conduit à un meilleur gain de survie à long terme que des essais traditionnels de grande taille, conçus pour maintenir des critères strictes avec un effet attendu faible.

La Figure 3.10 présente la distribution a posteriori du logarithme du Hazard Ratio suivant une loi normale (i) lorsqu'aucune donnée historique n'est incorporée ($\alpha_0 = 0$ et $\omega = 0$) et de paramètre ($m = -0.239$, $s^2 = 0.076$), (ii) lorsqu'une certaine quantité de données historiques est incorporée ($\alpha_0 = 0.3$ et $\omega = 0.1$) et de paramètre ($m = -0.245$, $s^2 = 0.042$) et (iii) lorsque toutes les données historiques sont incorporées ($\alpha_0 = 1$ et $\omega = 1$) et de paramètre ($m = -0.240$, $s^2 = 0.010$).

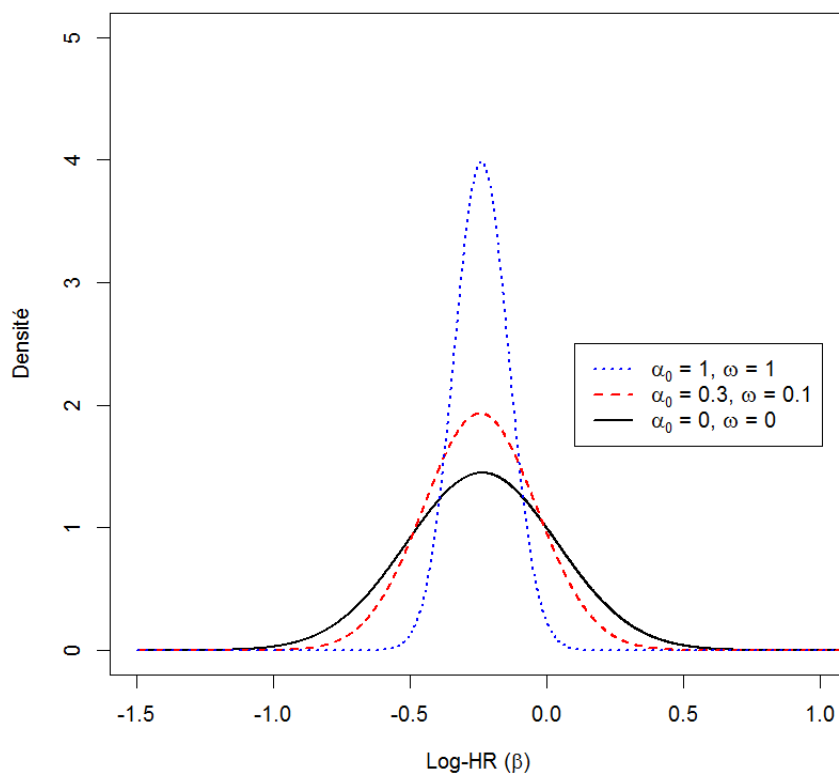


FIGURE 3.10 – Distribution a posteriori du Log(HR) selon différentes valeurs des paramètres de pondération dans le contexte du scénario S3

Dans le contexte du scénario S3 (données historiques et données du nouvel essai commensurables), cette figure représente les cas où aucune information historique n'est intégrée (courbe noire pleine), les données historiques sont pondérées selon la paramétrisation choisie pour l'essai Sarcome-13 (courbe rouge en tirets), et où les données historiques sont incorporées sans pondération (courbe bleue en pointillés).

Résultats pour une distribution exponentielle par morceaux

Des conclusions similaires sont constatées lorsque les données historiques et celles du nouvel essai suivent une distribution exponentielle par morceaux et sont analysées en utilisant un modèle Bayésien exponentiel par morceaux (Tableaux F4 à F6 de l'Annexe F).

3.4.4 Impact de l'inclusion de données historiques en cas de conflit en matière de distributions de survie

Le Tableau 3.4 résume l'impact de l'incorporation des données contrôles historiques uniquement (en fixant $\omega = 0$ et en faisant varier α_0), lorsque les données historiques suivent une distribution de Weibull et que les données du nouvel essai suivent une distribution exponentielle par morceaux (Scénario S27, effet du traitement du nouvel essai égal à l'effet du traitement historique, avec HR = 0.786).

Les données simulées de ce scénario ont été analysées en utilisant soit un modèle de Weibull (W/P/W), soit un modèle exponentiel par morceaux (W/P/P). Dans l'ensemble, le choix du modèle d'analyse a eu un impact limité.

Cela peut s'expliquer par le fait que ces jeux de données historiques étaient, par construction, relativement proches même s'ils provenaient de distributions de survie différentes (Figure 3.3). Cependant, dans le scénario S27, le biais dans l'estimation de l'effet du traitement est légèrement moins important si le modèle d'analyse est cohérent avec la distribution des données du nouvel essai (W/P/P). En effet, lorsque $\alpha_0 = 1$, le biais est de -0.070 dans la configuration W/P/P, contre -0.088 dans la configuration W/P/W, où le modèle d'analyse est cohérent avec la distribution des données historiques mais pas avec la distribution des données du nouvel essai.

Comparé au scénario S3, où il n'y a pas de conflit entre la distribution de survie des données historiques et celle du nouvel essai, lorsque les données sont analysées avec un modèle de Weibull (W/W/W), l'incorporation de données historiques en cas de conflit en matière de distribution de survie conduit à une augmentation du biais et à une perte d'exactitude globale de l'estimation.

3.4. Résultats

Des résultats similaires sont observés pour les scénarios S25, S26, S28 pour lesquels l'effet du traitement du nouvel essai diffère des données historiques (Détails des résultats disponibles dans les Tableaux H1 à H2 de l'Annexe H).

α_0	W/P/W				W/P/P				W/W/W			
	Biais	SD	RMSE	Puissance	Biais	SD	RMSE	Puissance	Biais	SD	RMSE	Puissance
0	-0.0142	0.292	0.298	0.358	-0.0087	0.287	0.292	0.345	0.0022	0.275	0.275	0.359
0.3	-0.0541	0.258	0.255	0.447	-0.0366	0.256	0.248	0.416	-0.0070	0.245	0.231	0.396
0.6	-0.0736	0.245	0.248	0.505	-0.0544	0.243	0.240	0.464	-0.0111	0.233	0.218	0.418
1	-0.0883	0.235	0.247	0.554	-0.0702	0.235	0.239	0.518	-0.0141	0.223	0.212	0.448

TABLEAU 3.4 – Impact d'un conflit en matière de distribution de survie entre les données historiques et actuelles pour le scénario S27 en fonction de α_0 ^a

^a Résultats du scénario S27, où une distribution de Weibull est utilisée pour générer les données historiques contrôles, et où une distribution exponentielle par morceaux est utilisée pour générer les données du nouvel essai avec HR = 0.786. Ces données sont analysées, avec $\omega = 0$, soit avec un modèle de Weibull (W/P/W), soit avec un modèle exponentiel par morceaux (W/P/P). Ces résultats sont comparés au scénario S3, pris comme référence, et défini par des données commensurables suivant une distribution de Weibull, et analysées avec un modèle de Weibull (W/W/W).

SD : Ecart-type (*Standard Deviation*)

RMSE : Racine carrée de l'erreur quadratique moyenne (*Root Mean Square Error*)

Lorsque les données historiques suivent une distribution exponentielle par morceaux et que les nouvelles données suivent une distribution de Weibull, les deux types de modèles (Weibull et exponentiel par morceaux) conduisent à des résultats très similaires (Tableau H3 pour P/W/P et Tableau H4 pour P/W/W de l'Annexe H).

3.4.5 Simulations intégrant des données réelles historiques

Dans la version finale du protocole de l'essai Sarcome-13, de légers changements relatifs aux critères d'éligibilité apportés, par rapport à la version préliminaire utilisée pour cette étude de simulations. Cela a eu pour conséquence de modifier l'échantillon de patients historique issu de l'essai OS2006 (de 165 patients et 73 événements, à 256 patients et 116 événements), et d'augmenter l'échantillon de patients planifié pour l'essai Sarcome-13 (de 105 initialement à 126 patients dans la version finale du protocole).

Le découpage du temps du modèle exponentiel par morceaux a également été redéfini de 1 et 2 ans initialement, à 0.3 et 1.6 ans.

Par conséquent, nous avons décidé de réaliser de nouvelles simulations en considérant un petit ensemble de scénarios (voir Tableau 3.5) pour explorer le voisinage du couple de paramètres précédemment défini ($\alpha_0 = 0.3$, et $\omega = 0.1$).

Nous avons considéré les valeurs suivantes :

- $\alpha_0 \in \{0, 0.1, 0.2, 0.3, 0.4\}$
- $\omega \in \{0, 0.05, 0.1, 0.15\}$

Les scénarios évalués sont équivalents aux scénarios S13-S24, excepté que les données historiques utilisées sont les données réelles SARC-OS mises à jour (256 patients). Ce nouvel ensemble de scénarios seront appelé S33-S44.

Pour chaque scénario, 5 000 essais ont été simulés avec un échantillon de 126 patients et un taux de censure de 5%. Cette étude de simulations a été jointe au protocole et à l'article soumis au journal *BMJ Open*.

3.4. Résultats

Scénarios	Données historiques contrôles	Génération des nouvelles données		
		Distribution de survie ^a	Paramètres	
			Bras contrôle	Effet traitement ^b
S33			Nul	
S34			Décevant	
S35			Historique	
S36			Espéré	
S37			Nul	
S38	Données Sarc-OS	Exponentielle par morceaux	Conflit	Décevant
S39			négatif	Historique
S40				Espéré
S41				Nul
S42			Conflit	Décevant
S43			positif	Historique
S44				Espéré

TABLEAU 3.5 – **Résumé des 12 scénarios considérant les données réelles SARC-OS**

^a Distribution de survie utilisée pour générer les données du nouvel essai

^b Les effets nul, décevant, historique et espéré correspondent à un hazard ratio de 1, 0.886, 0.786, et 0.55 dans le nouvel essai, respectivement

Lorsque $\alpha_0 = 0$, nous recommandons de fixer $\omega = 0.1$, comme déterminé précédemment. Cette valeur de ω conduit à une erreur de type I de 22.2% (scénario S33) jugée acceptable dans le contexte de l'essai Sarcome-13, tandis qu'elle permet un gain substantiel de puissance en augmentant ω de 0 à 0.1 (de 38.1% à 56.9% dans le scénario S35 par exemple), et une meilleure exactitude globale (la RMSE diminue de 0.259 à 0.196 dans le scénario S35).

En utilisant la configuration $\alpha_0 = 0$ et $\omega = 0$ comme référence, nous avons conclu que l'incorporation de données historiques en fixant $\alpha_0 = 0.3$ et $\omega = 0.1$ est un compromis acceptable entre augmentation de la puissance et de la précision, et augmentation de l'erreur de type I et du biais. En effet, ce couple de valeurs permet d'augmenter la puissance de l'essai de 22.7% à 42.1% dans le scénario "effet décevant" (S34) ; de 38.1% à 63.8% dans le scénario "effet historique" (S35) ; et de 82.4% à 97.8% dans le scénario "effet espéré" (S36).

La configuration $\alpha_0 = 0.3$ et $\omega = 0.1$ conduit également à une erreur de type I acceptable, dans le contexte de l'essai Sarcome-13, de 21% (scénario S33) lorsque les données contrôles historiques et les données contrôles du nouvel essai sont commensurables. De plus, lorsque ce n'est pas le cas (scénarios S37-S44), les risques de biais, de diminution de la puissance, ou d'inflation de l'erreur de type I sont acceptables en fixant ces valeurs pour α_0 et ω . Le résultats de cette nouvelle étude de simulations sont donc cohérents avec ceux retrouvés dans l'étude de simulations plus générale (Détails des résultats disponibles dans les Tableaux I1 à I3 de l'Annexe I).

3.5 Discussion

Dans ce travail, nous proposons l'utilisation d'un *power prior* (avec un paramètre de pondération fixe) et d'un *mixture prior* pour incorporer simultanément des données historiques contrôles individuelles et des données agrégées sur l'effet du traitement dans l'analyse Bayésienne de la survie d'un nouvel essai.

Les caractéristiques opératoires de l'essai variaient selon le poids attribué à chaque source d'information historique, la variance des composantes informative et vague du *mixture prior*, et le niveau de commensurabilité entre les données historiques et celles du nouvel essai. À noter que le choix du paramètre de pondération du *mixture prior* est conditionné par le choix de la variance de la composante vague de cette distribution a priori.

L'incorporation de données historiques sur le bras contrôle dans l'analyse semble n'avoir que peu d'impact sur la puissance lorsque les données historiques et actuelles sont commensurables. Cependant, en cas de non-commensurabilité, le risque de biais dans l'estimation de l'effet du traitement est important. L'intégration de données individuelles sur le bras contrôle pour informer l'estimation de l'effet du traitement d'un essai reste donc un choix discutable. Les gains de puissance et de précision possibles en incorporant ces données lorsqu'elles sont commensurables avec les données du nouvel essai, doivent

donc être mis en balance avec le risque de biaiser les estimations, et de perdre en puissance si les données ne sont pas commensurables.

Au terme de notre étude de simulations, nous avons identifié empiriquement les valeurs pour α_0 et ω qui constituaient un compromis raisonnable entre puissance, biais, et précision, pour une étude de petite taille ayant un cadre similaire à l'essai Sarcome-13.

Ce travail de calibration par une étude de simulations nous semble nécessaire. Nous ne pouvons cependant ignorer ni la complexité technique de cette étape préalable à la soumission du protocole, ni le temps de calcul correspondant. Nous avons finalement dû avoir recours à un serveur de 125 coeurs, en parallélisant les analyses avec le logiciel R (package R Parallel [162]) pour réaliser les simulations dans un temps raisonnable. Dans ces conditions, avec la réalisation d'une seule chaîne pour 20 000 itérations, l'analyse de 28 paires (α_0, ω) pour un scénario donné, évalué au travers de 5 000 essais simulés, nécessite deux heures si l'analyse repose sur un modèle de Weibull Bayésien et plus de dix heures pour un modèle exponentiel par morceaux à 3 paramètres.

Outre l'évaluation de l'impact du poids alloué aux données historiques, nous avons également évalué l'impact de l'incertitude du modèle s'ajustant le mieux aux données (lorsque les données individuelles contrôles historiques utilisées pour déterminer le modèle d'analyse s'ajustent mal avec les données contrôles du nouvel essai en matière de forme de distribution de survie).

L'intégration de données individuelles contrôles historiques nécessite une modélisation de la survie. La définition du type de modèle de survie peut être difficile en raison de l'incertitude qui entoure souvent la forme de la distribution dans les pathologies rares. Dans notre étude de simulations, quel que soit le scénario, choisir un modèle exponentiel par morceaux par rapport à un modèle de Weibull pour l'analyse Bayésienne du nouvel essai, n'apportait pas d'avantages en matière de biais, de précision, ou de puissance. Ces résultats peuvent s'expliquer par la similitude entre les distributions des jeux de données historiques hypothétiques et celle du jeu de données généré pour représenter le nouvel essai dans le contexte de l'essai Sarcome-13.

Les résultats présentés dans ce chapitre ont été obtenus à partir d'une approche Bayésienne utilisant deux distributions a priori différentes (*power prior* et *mixture prior*) afin d'incorporer deux types de données historiques différents (individuelles et agrégées). Nous aurions pu utiliser des distributions a priori similaires pour les deux types de données mais notre approche pragmatique a deux principaux avantages :

- Appliquer un *mixture prior* aux données agrégées permet d'avoir une composante robuste à même de mieux répondre au conflit entre la distribution a priori et les nouvelles données.
- Appliquer un *power prior* (avec un paramètre de pondération fixe) aux données historiques individuelles est moins complexe, car un *mixture prior* nécessiterait une hypothèse asymptotique sur la forme de la distribution a posteriori jointe des paramètres du modèle de survie du groupe contrôle. Étant donné qu'il y aura souvent plusieurs paramètres caractérisant la survie dans le groupe contrôle, le *power prior* est une manière plus parcimonieuse d'intégrer ces données.

L'évaluation de l'incorporation de données historiques du même type dans l'analyse d'un nouvel essai a déjà été réalisée dans le cadre d'un critère binaire [90, 107, 152, 153, 167, 168] et de survie [169], avec des résultats similaires à ceux présentés ici.

Dans le cadre d'un critère binaire, Cuffe [167] alerte sur les risques de diminuer la puissance, et de biaiser l'estimation du taux de réponse, en incorporant des données historiques contrôles si celles-ci ne sont pas commensurables avec les nouvelles données, ce qui est en accord avec nos résultats. De plus, il souligne que si nous voulons assurer le contrôle de l'erreur de type I, en prenant en compte un conflit possible entre la distribution a priori et les données, nous risquons de ne pas augmenter la puissance ou même de la réduire.

Plus récemment, Li et al. [152] présentent une distribution a priori prédictive méta-analytique empirique pour mieux ajuster le poids des données historiques en fonction du degré de commensurabilité entre la distribution a priori et les nouvelles

données. Ils comparent leur méthode au *meta-analytic predictive prior*, et sa version robuste présentée par Schmidli et al. [90]. Ils montrent que leur méthode permet de mieux contrôler l'erreur de type I en cas d'hétérogénéité entre les données historiques et le nouvel essai. Ils soulignent également la difficulté de déterminer la valeur du poids ω dans le *meta-analytic predictive prior* puisque qu'il serait basé sur la confiance des investigateurs dans la pertinence des données historiques.

Lorsqu'un seul essai historique est disponible, Gravestock et al. [153] proposent une approche empirique de Bayes pour construire le *power prior*, afin de répondre de façon dynamique au conflit potentiel entre la distribution a priori et le nouvel essai. Cette approche a de bonnes performances par rapport à une approche de Bayes complète, ou une approche avec un paramètre de pondération α_0 fixe dont le choix n'est pas simple.

Dans le contexte d'un critère censuré, van Rosmalen et al. [169] comparent différentes méthodes pour inclure les données individuelles du bras contrôle de plusieurs essais historiques. Ils démontrent que prendre en compte l'hétérogénéité entre l'essai historique et actuel est nécessaire pour pleinement tirer parti des données historiques. Plus généralement, Neuenschwander et al. [88], Gsteiger et al. [89], et Schmidli et al. [90] adoptent un modèle méta-analytique pour décrire les différences entre les essais historique et actuel, pour un paramètre d'intérêt. La distribution a priori de ce paramètre est dérivée en supposant l'échangeabilité des données.

Comme indiqué dans la Section 3.2, nous avons choisi de travailler avec des paramètres α_0 et ω fixes et n'avons pas tenté de prendre en compte de façon dynamique les différences observées entre les essais historiques et le nouvel essai pour les paramètres d'intérêt. Il était possible d'envisager des *priors* adaptatifs avec des paramètres sélectionnés dynamiquement lors des analyses en fonction de la commensurabilité des données de l'essai avec les données historiques. Cependant, dans le contexte des pathologies rares, les différences observées entre les données historiques et les données de l'essai actuel peuvent être dues à des variations aléatoires plus qu'à de vraies différences, risquant à tort de conduire à une remise en question de la commensurabilité.

Concernant le *mixture prior* utilisé pour le logarithme du hazard ratio, sélectionner les valeurs les plus appropriées pour ω , et pour la variance de la composante vague, n'est pas simple. Les deux valeurs auront une incidence sur les caractéristiques opératoires de l'essai.

Dans ce chapitre, toutes les simulations ont été réalisées en définissant la composante vague comme étant une distribution normale de moyenne zéro et de variance 10. Mutsvari et al.[91] soulignent l'importance du choix de la variance de cette composante afin d'assurer une pondération adéquate de l'information historique en cas de conflit entre la distribution a priori et les nouvelles données. Si cette variance est trop importante, le *mixture prior* aura des queues de distribution très lourdes, plaçant la masse de la distribution a priori sur des valeurs de l'effet du traitement très grandes et improbables. Cela entraînerait alors une faible pondération des données historiques, même s'il existe un conflit évident entre la distribution a priori et les nouvelles données. C'est pourquoi, après avoir réalisé des simulations, nous avons choisi une variance de 10 permettant une pondération plus progressive.

Nous n'avons pas évalué l'impact de différents ratios entre les nombres de patients contrôles historiques et actuels, sur les caractéristiques opératoires de l'essai. Puisque l'impact de l'inclusion de données individuelles contrôles historiques est faible, même lorsque le ratio est proche de 3 (comme dans le cadre de ce chapitre; c'est-à-dire 52 patients dans le nouvel essai *versus* 165 patients dans l'essai historique), nous pouvons extrapoler pour conclure que varier ce ratio aurait eu peu d'impact sur les propriétés de l'essai.

De plus, nous n'avons pas évalué la manière dont le ratio de randomisation du nouvel essai impacterait l'analyse si l'on recrutait plus de patients dans le bras expérimental. Cependant, dans le contexte de l'essai Sarcome-13, les investigateurs ont préféré maintenir un ratio de randomisation 1:1, ce qui est généralement choisi dans les essais portant sur des pathologies rares.

Afin d'évaluer l'apport de l'incorporation des données historiques dans l'analyse de l'essai par rapport à une analyse fréquentiste classique, il nous a semblé nécessaire de spécifier une règle de décision pour l'analyse Bayésienne et que cette règle soit proche de celle de l'analyse fréquentiste ($P(HR < 1)$). Il est cependant prévu de décrire la distribution a posteriori de l'effet du traitement et ainsi la probabilité d'avoir une certaine quantité d'effet (et pas seulement un effet non nul). Ceci est présenté dans la publication soumise rapportant le protocole de l'essai.

Suivant les recommandations de la FDA dans le cadre des dispositifs médicaux [79], nous avons de plus proposé une analyse "fréquentiste" des caractéristiques opératoires du design proposé, ce qui peut surprendre dans le contexte Bayésien. En effet la FDA recommande d'évaluer notamment l'erreur de type I, et la puissance. De plus, utiliser le même critère d'évaluation nous permet également de comparer le design proposé à l'approche classique fréquentiste. Pour autant, dans le cadre Bayésien, d'autres métriques d'évaluation ont été proposées telles que la probabilité de succès calculée comme une moyenne pondérée d'une conclusion positive pour différentes valeurs possibles de l'effet traitement.

Dans ce travail, notre préoccupation principale était d'intégrer des données historiques pour contribuer à l'estimation de l'effet d'un traitement expérimental. Nous avons choisi une certaine approche Bayésienne. Comme évoqué précédemment, d'autres approches Bayésiennes étaient également possibles. Par ailleurs, une méta-analyse classique permettrait également d'évaluer l'effet du traitement expérimental sur la base de l'essai actuel et des essais passés. Cependant, la finalité d'une méta-analyse est de fournir une estimation globale de l'effet du traitement contrairement à l'approche que nous proposons centrée sur l'estimation de l'effet du traitement sur la base des données de l'essai actuel, à la lumière des données historiques. De plus, le poids apporté aux données historiques est pondéré négativement dans l'approche Bayésienne. Une comparaison formelle des deux approches pourrait être envisagée.

Chapitre 3. Proposition d'un design Bayésien avec incorporation de données historiques individuelles et agrégées pour un essai clinique

Bien que cette étude de simulations soit basée sur l'essai Sarcome-13, nous avons considéré différents scénarios de simulations afin de permettre à nos conclusions d'être informatives pour les statisticiens souhaitant utiliser une approche Bayésienne, dans un essai considérant une pathologie rare.

— Chapitre 4 —

Expression des résultats d'une analyse Bayésienne de la survie d'un essai clinique randomisé

4.1 Introduction

L'objectif de ce 3ème axe est de présenter l'apport d'une analyse Bayésienne par rapport à une analyse classique fréquentiste dans l'analyse finale d'un essai randomisé avec un critère censuré. En effet, il est fréquemment rapporté que l'utilisation de l'approche Bayésienne permet d'enrichir l'interprétation des résultats, même en l'absence de données historiques pertinentes, fournissant une description complète de la distribution de l'effet d'un traitement, et une interprétation des résultats plus facile pour les cliniciens [96, 170, 171]. Enfin, cette approche permet de choisir d'autres règles de décision que le rejet de l'hypothèse nulle.

On peut ainsi évaluer un traitement à l'aune de la probabilité qu'il soit associé à un effet cliniquement pertinent, en considérant les recommandations des experts pour la définition des seuils.

Pour mémoire, le groupe d'experts de l'*European Society for Medical Oncology* (ESMO), recommande de comparer la borne inférieure de l'intervalle de confiance à 95% (IC95%) à une valeur seuil [7].

Nous ne considérerons ici que l'approche Bayésienne dans le cadre de l'analyse finale d'un essai. L'effet du traitement expérimental par rapport au traitement contrôle sera caractérisé par le rapport des risques instantanés, estimé par le hazard ratio.

4.2 Matériels et méthodes

4.2.1 Applications à des essais randomisés réels en oncologie

Afin d'illustrer l'apport de l'approche Bayésienne dans l'expression des résultats, nous présentons trois exemples d'application d'une analyse Bayésienne à des essais réels oncologiques de supériorité, ayant pour critère principal un critère censuré, et dont l'analyse principale a été une analyse classique fréquentiste, et pour lesquels nous disposons des données individuelles.

Nous avons choisi de travailler sur les données de deux essais de petite taille dans des pathologies rares (essai EE99-R2Loc dans les sarcomes d'Ewing [172] et essai HERBY dans les gliomes de l'enfant [173]) et d'un essai de grande taille (IALT dans le cancer du poumon non à petites cellules [174]). Ce dernier essai a été choisi pour illustrer les différences entre l'approche fréquentiste classique et l'approche Bayésienne, en fonction de la taille de l'essai.

Les résultats des analyses publiées sont présentés dans le Tableau 4.1. Deux essais ont conclu au bénéfice significatif du traitement évalué sur le critère de jugement principal (essais EE99-R2Loc et IALT) contrairement au troisième (essai HERBY).

Essai EE99-R2Loc

EE99-R2Loc [172] est un essai de phase III randomisé [172], qui évaluait l'effet de la chimiothérapie de consolidation à haute dose par Busulfan-Melphalan (groupe expérimental) en comparaison à la chimiothérapie standard (groupe contrôle), chez des enfants et des adultes de moins de 50 ans, atteints d'un sarcome d'Ewing localisé à haut risque de rechute. Le critère principal de cet essai était la survie sans événement (EFS), la survie globale étant estimée en critère secondaire.

Le design de cet essai a été construit afin d'assurer une puissance de 80% pour une amélioration de l'EFS à 3 ans de 25% en différence absolue (EFS à 3 ans attendue, 70% *versus* 55%, équivalent à un HR = 0.60 sous l'hypothèse de proportionnalité des risques) avec un test du logrank réalisé au seuil alpha bilatéral = 5%. La taille de l'échantillon initialement prévue était de 328 patients (124 événements). À noter que l'effet du traitement considéré dans l'hypothèse alternative du calcul de nombre de sujets nécessaire était sciemment très optimiste pour des raisons pragmatiques compte tenu de la rareté de la pathologie étudiée. L'analyse principale de l'essai a été réalisée par un modèle de Cox univarié et non stratifié.

L'essai a été interrompu "prématurément" sans avoir atteint le nombre de patients initialement prévu (N = 328), en raison des difficultés de recrutement. La décision d'arrêt a été prise indépendamment de l'effet du traitement observé. L'analyse publiée est donc considérée comme l'analyse finale et non comme une analyse intermédiaire.

L'essai a inclus 240 patients (118 dans le groupe contrôle, et 122 dans le groupe expérimental) entre février 2000 et décembre 2015. Les auteurs ont conclu à un effet positif significatif de la chimiothérapie de consolidation comparée à la chimiothérapie standard sur la survie sans événement des patients, avec un HR = 0.64 [IC95% : 0.43-0.95] p = 0.026.

Essai IALT

L'essai IALT [174] est un essai de phase III randomisé qui évaluait le bénéfice apporté par la chimiothérapie adjuvante incluant du Cisplatine (groupe expérimental) en comparaison à un bras observationnel (groupe contrôle), chez des patients âgés de 18 à 75 ans atteints d'un cancer du poumon non à petites cellules. Le critère principal de cet essai était la survie globale (OS).

L'essai a été construit de façon à assurer une puissance de 90% pour une amélioration de la survie à 5 ans de 5% en différence absolue (OS à 5 ans attendue, 50% *versus* 55%, équivalent à un HR = 0.85 sous l'hypothèse de proportionnalité des risques) avec un test du logrank réalisé au seuil alpha bilatéral = 5%. La taille de l'échantillon initialement prévue était de 3 300 patients (1 600 décès). L'analyse principale de l'essai a été réalisée par un modèle de Cox ajusté sur les variables prises en compte dans la randomisation et stratifié sur le centre.

Il a été décidé d'arrêter le recrutement des patients précocement à cause d'un ralentissement important du taux de recrutement. Là encore, l'analyse réalisée a alors été considérée comme l'analyse finale et non comme une analyse intermédiaire.

L'essai a inclus 1 867 patients (932 dans le groupe contrôle, et 935 dans le groupe expérimental) entre février 1995 et décembre 2000.

Les auteurs ont conclu à l'effet positif de la chimiothérapie adjuvante sur la survie globale des patients, comparé à l'observation, avec un HR = 0.86 [IC95% : 0.76-0.98] p = 0.023.

Essai HERBY

HERBY est un essai de phase II randomisé [173], qui évaluait l'efficacité du Bevacizumab en association à la radiothérapie + Temozolomide (groupe expérimental) *versus* radiothérapie + Temozolomide seul (groupe contrôle), chez des patients âgés de 3 à 18 ans atteints d'un gliome de haut-grade. Le critère principal de cet essai était la survie sans événement, la survie globale étant évaluée en critère secondaire.

4.2. Matériels et méthodes

Il était prévu de recruter 120 patients (89 événements) dans cet essai afin d'assurer une puissance de 80% pour une amélioration de l'EFS à 1 an de 20% en différence absolue (EFS à 1 an attendu, 62% *versus* 42%, équivalent à un HR = 0.55 sous l'hypothèse de proportionnalité des risques), avec un test du logrank réalisé au seuil alpha bilatéral = 5%. L'analyse de l'essai a été réalisée à l'aide d'un modèle de Cox stratifié sur l'âge, la qualité de la résection et le grade de la tumeur.

L'essai a inclus 121 patients (59 dans le groupe contrôle, et 62 dans le groupe expérimental) entre octobre 2011 et février 2015. L'analyse présentée correspond à l'analyse finale prévue 1 an après l'inclusion du dernier patient, telle que prévue dans le protocole.

Les auteurs ont conclu à l'absence d'efficacité du Bevacizumab associé à la radiothérapie + Temozolomide comparé à la radiothérapie + Temozolomide seuls, avec un HR = 1.44 [IC95% : 0.90-2.30] p = 0.132.

Les résultats de l'analyse sur le critère de jugement principal des trois essais sont résumés dans le Tableau 4.1.

Disposant des données individuelles pour ces essais, nous avons pu reproduire les analyses et ainsi récupérer les valeurs précises des estimations dans les modèles correspondants (avec ajustement et stratification le cas échéant).

Essais	Taille d'effet attendu (HR) ^a	Nombre de patients recrutés	Nombre d'évènements ou de décès observés	Hazard Ratio observé [IC95%]	p-valeur	Log(HR) (se) ^b
EE99-R2Loc	0.60	240	105	0.64 [0.43-0.95]	0.026	-0.448 (0.198)
IALT	0.85	1867	973	0.86 [0.76-0.98]	0.023	-0.148 (0.065)
HERBY	0.55	121	75	1.44 [0.90-2.30]	0.132	0.361 (0.240)

TABLEAU 4.1 – Résultats fréquentistes des essais réanalysés

^a Hazard ratio "espéré", utilisé pour définir la puissance de l'essai (hypothèse alternative, H₁) dans le calcul du nombre de sujets nécessaire

^b Logarithme du Hazard ratio et son erreur standard estimé (dans le modèle correspondant à l'analyse publiée avec ajustement et/ou stratification le cas échéant)

4.2.2 Plan de l'analyse Bayésienne

L'effet du traitement sur le critère de jugement principal étant caractérisé par le rapport des risques instantanés (Hazard Ratio), on suppose que le logarithme du HR ($\text{Log}(\text{HR})$) suit une loi normale [47].

Dans un premier temps, une distribution a priori non-informative a été utilisée. Nous avons supposé que le $\text{Log}(\text{HR})$ suivait une loi normale de moyenne nulle et d'écart-type 100. Ceci signifie qu'aucune contribution n'est apportée par la distribution a priori, la valeur choisie pour la moyenne n'a donc aucun impact. Les résultats de cette analyse coïncideront ainsi avec l'analyse fréquentiste de chaque essai; cependant, l'approche Bayésienne nous permettra de représenter la distribution de l'effet du traitement.

Nous avons par la suite réalisé une analyse en intégrant deux distributions a priori supplémentaires archétypales pour le $\text{Log}(\text{HR})$:

- Une distribution a priori enthousiaste centrée sur la "taille d'effet attendu" (utilisée pour le calcul de la puissance de l'essai), représentant l'optimisme face à un certain effet du traitement
- Une distribution a priori sceptique centrée en zéro (absence d'effet du traitement), représentant le pessimisme face à un quelconque effet du traitement

L'écart-type de ces deux distributions a priori a été défini selon la méthode décrite par Freedman et al. [100], ainsi :

$$\sigma_0 = \frac{|\text{Ln}(\text{HR}_1)|}{z_{1-\gamma}} \quad (4.1)$$

où HR_1 est le HR considéré pour l'hypothèse alternative (H_1), utilisée lors du calcul de la puissance de l'essai, et $z_{1-\gamma}$ est le quantile d'ordre $1 - \gamma$ d'une loi normale (égal à 1.645 pour $\gamma = 5\%$). La quantité γ est définie pour les trois essais à 5% et représente la probabilité d'excéder HR_1 .

4.2. Matériels et méthodes

Cet écart-type peut être traduit en nombre de pseudo-événements (N_0) afin de mieux rendre compte de la quantité d'information contenue dans la distribution a priori. Ce nombre de pseudo-événements est défini comme :

$$d_0 = \frac{z_{1-\gamma}^2}{(z_{1-\alpha/2} + z_{1-\beta})^2} \times d_1 \quad (4.2)$$

où α et β représentent l'erreur de type I et l'erreur de type II définis dans le protocole de l'essai, et d_1 est le nombre d'événements nécessaire calculé pour l'essai.

Le Tableau 4.2 présente les différentes distributions a priori utilisées pour chaque essai.

Essais	Nombre d'événements attendus (d_1)	Distributions a priori du Log(HR) (en supposant une loi normale N(moyenne, écart-type))		
		Non-informatif	Sceptique	Enthousiaste
EE99-R2Loc	124	N(0, 100) $d_0 < 0.001$	N(0, 0.311) $d_0 = 43$	N(-0.511, 0.311) $d_0 = 43$
IALT	1 600	N(0, 100) $d_0 < 0.001$	N(0, 0.099) $d_0 = 412$	N(-0.163, 0.099) $d_0 = 412$
HERBY	89	N(0, 100) $d_0 < 0.001$	N(0, 0.364) $d_0 = 31$	N(-0.598, 0.364) $d_0 = 31$

TABLEAU 4.2 – Distributions a priori du logarithme du hazard ratio (Log(HR)) pour chaque essai

Log(HR) : Logarithme du Hazard Ratio

d_0 : Nombre de pseudo-événements calculé à partir de la formule 4.2 issue de l'article de Freedman et al. [100]

Les distributions a posteriori de l'effet du traitement Log(HR) pour chaque essai, ont été estimées par un modèle de Cox Bayésien en utilisant un processus de type Monte Carlo par chaînes de Markov (*Markov Chain Monte Carlo*; *MCMC*). Nous avons utilisé une chaîne de Markov et l'algorithme de Metropolis-Hastings [52] pour 25 000 itérations, dont 5 000 burn-in, pour les inférences a posteriori. La convergence des distributions a été évaluée graphiquement, et à l'aide du test diagnostique de Geweke [53].

La procédure PROC PHREG du Logiciel SAS a été utilisée pour réaliser les analyses finales selon l'approche fréquentiste et l'approche Bayésienne.

Nous avons envisagé initialement une approche hybride (utilisant les lois conjuguées). En effet, la distribution a posteriori du Log(HR) peut s'exprimer de la manière suivante en utilisant une distribution a priori conjuguée :

$$N\left(\frac{d_0\mu_0 + d\hat{\mu}}{d_0 + d}, \sqrt{\frac{4}{(d_0 + d)}}\right)$$

où μ_0 est la moyenne (Log(HR)) a priori, $d_0 = 4/\sigma_0$ est le nombre d'événements a priori et où σ_0 est l'écart-type a priori, d est le nombre d'événements observés, et $\hat{\mu}$ est le Log(HR) estimé par maximisation de la vraisemblance partielle (modèle de Cox).

Cependant, cette approche n'était pas possible ici car les analyses des essais IALT et HERBY étaient stratifiées et/ou ajustées. L'approximation nécessaire de la variance par l'approche proposée par Schoenfeld [163] ($\sigma^2 = 4/d$) ne donnait pas des résultats cohérents. Nous avons donc utilisé un modèle de Cox Bayésien avec une estimation par un processus MCMC pour estimer la loi a posteriori empirique et ses paramètres (moyenne et écart-type).

Nous avons ensuite pu calculer la probabilité d'observer un effet du traitement supérieur ou inférieur à différents seuils. Les valeurs seuils considérées correspondent au rejet de l'hypothèse nulle, aux hypothèses alternatives considérées dans le protocole des essais étudiés, et enfin à une valeur admise comme "cliniquement pertinente" (Borne HR = 0.7 comme proposé en Section 1.2.4 sur la base des recommandations d'experts (ESMO) [7]). Nous avons considéré que le Log(HR) suivait une distribution normale avec les paramètres (moyenne et écart-type) estimés précédemment à partir de la loi empirique afin de représenter graphiquement la distribution a priori, la vraisemblance et la distribution a posteriori.

4.3 Résultats

4.3.1 Analyse Bayésienne avec distribution a priori non-informative

L'essai EE99-R2Loc a conclu à l'efficacité du traitement comparé à la chimiothérapie standard avec un HR = 0.64 IC95%[0.43-0.95] ($p = 0.026$). La moyenne de la distribution a posteriori empirique de l'effet du traitement, utilisant une distribution a priori non-informative (Figure 4.1), est égale à -0.450, équivalent à un HR = 0.64 avec une erreur standard (se) de 0.198, permettant de définir l'intervalle de crédibilité à 95% suivant ICr95%[-0.842 ; -0.061] pour le Log(HR), équivalent à un ICr95%[0.43;0.94] pour le HR, ce qui est cohérent avec le résultat de l'approche fréquentiste.

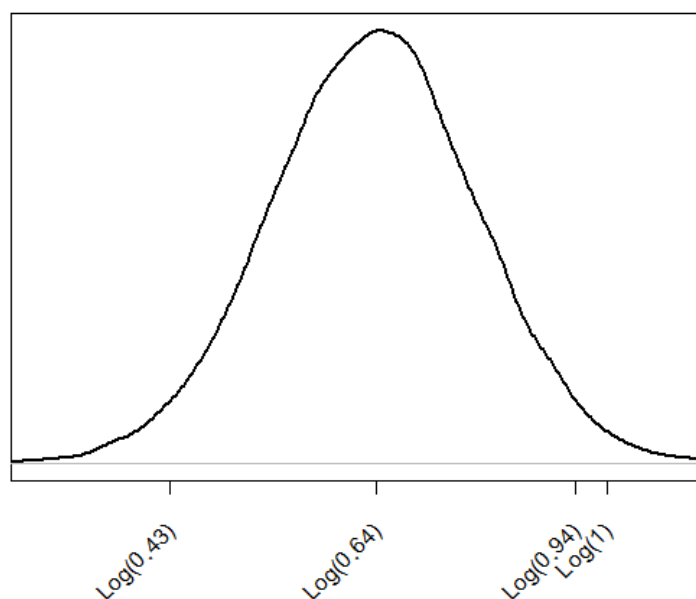


FIGURE 4.1 – Distribution a posteriori empirique du Log(HR) de l'essai EE99-R2Loc, estimé par un modèle de Cox Bayésien

Distribution a posteriori empirique de l'effet du traitement (Log(HR)) estimée par un modèle de Cox Bayésien au travers d'un processus MCMC, à partir des données de l'essai EE99-R2Loc et d'une distribution a priori non-informative ($\text{Log}(HR) \sim N(0,100)$). Sur l'axe des abscisses sont représentés les bornes de l'intervalle de crédibilité à 95% [Log(0.43);Log(0.94)], la médiane (Log(0.64)), et la borne d'absence d'effet traitement (Log(1)).

La Figure 4.2 représente la distribution d'une loi normale de moyenne ($= -0.450$) et d'écart-type ($= 0.198$) estimés par l'analyse Bayésienne. À partir de cette distribution, la probabilité a posteriori qu'il y ait un effet bénéfique du traitement, c'est-à-dire que l'effet du traitement ($\text{Log}(\text{HR})$) soit inférieur à zéro, est de 98.8% (aire grisée de la Figure 4.2, Tableau 4.3), ou de façon équivalente, le probabilité que le $\text{Log}(\text{HR})$ soit supérieur ou égal à 0 (donc qu'il n'existe aucun effet bénéfique) est de 1.2%. Cette probabilité est cohérente avec la p-valeur obtenue avec l'analyse fréquentiste qui, elle, correspond à la probabilité d'observer ce résultat sous l'hypothèse nulle d'absence d'effet du traitement ($0.012 \times 2 = 0.024$, proche de la p-valeur 0.026 du test bilatéral)

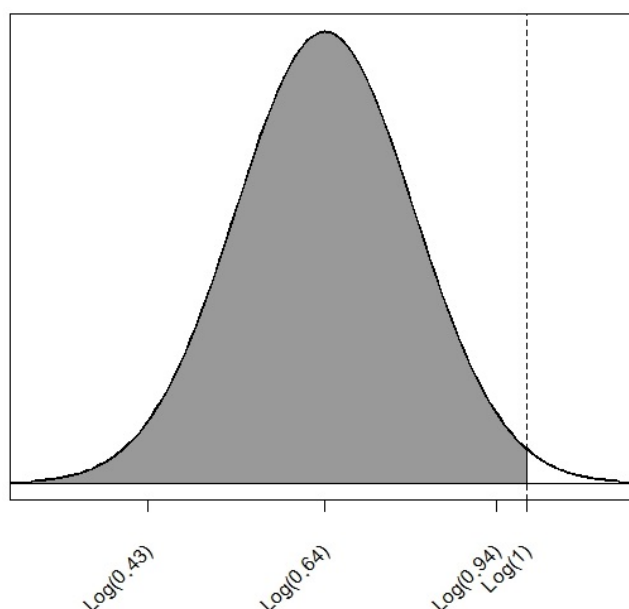


FIGURE 4.2 – Représentation de la probabilité a posteriori de l'effet du traitement $P(\text{Log}(\text{HR}) < 0$), dans l'essai EE99-R2Loc

Probabilité a posteriori de l'effet du traitement ($\text{Log}(\text{HR})$) estimée par l'analyse Bayésienne à partir des données de l'essai EE99-R2Loc (fonction de densité d'une loi normale de moyenne $= -0.450$ et écart-type $= 0.198$) et d'une distribution a priori non-informative ($\text{Log}(\text{HR}) \sim N(0, 100)$). Dans cet essai $P(\text{HR} < 1) = 0.989$, ce qui est cohérent avec la p-valeur de 0.026 de l'analyse fréquentiste ($0.012 \times 2 = 0.024$). Sur l'axe des abscisses sont représentés les bornes de l'intervalle de crédibilité à 95% [$\text{Log}(0.43)$; $\text{Log}(0.94)$], la médiane ($\text{Log}(0.64)$), et la borne d'absence d'effet traitement ($\text{Log}(1)$).

L'essai IALT a également conclu à l'efficacité du traitement évalué avec un HR = 0.86 IC95%[0.76-0.98] ($p = 0.023$). La moyenne de la distribution a posteriori de l'effet du traitement, estimée par l'analyse Bayésienne utilisant une distribution a priori non-informative (Figure 4.3), est égale à -0.148, équivalent à un HR = 0.86, avec une erreur standard $se = 0.065$, permettant d'estimer l'intervalle de crédibilité à 95% suivant ICr95%[-0.276;-0.020] pour le Log(HR), équivalent à un ICr95%[0.76;0.98] pour le HR.

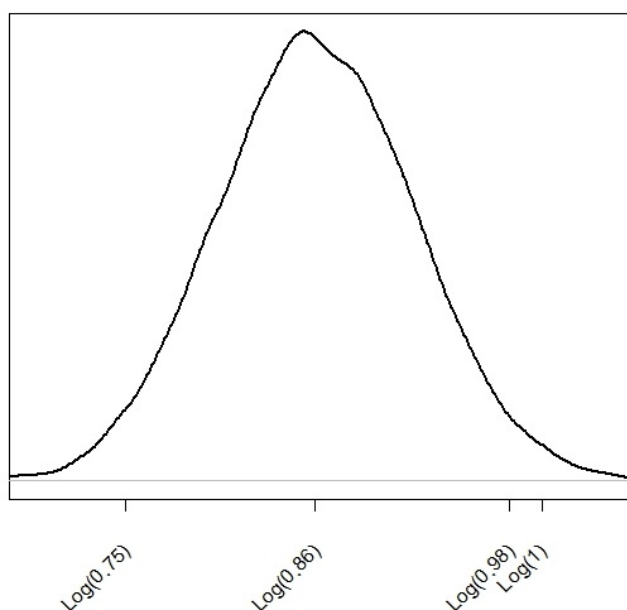


FIGURE 4.3 – Distribution a posteriori empirique du Log(HR) de l'essai IALT, estimé par un modèle de Cox Bayésien

Distribution a posteriori empirique de l'effet du traitement (Log(HR)) estimée par un modèle de Cox Bayésien au travers d'un processus MCMC, à partir des données de l'essai IALT et d'une distribution a priori non-informative ($Log(HR) \sim N(0,100)$). Sur l'axe des abscisses sont représentés les bornes de l'intervalle de crédibilité à 95% [Log(0.75);Log(0.98)], la médiane (Log(0.86)), et la borne d'absence d'effet traitement (Log(1)).

La Figure 4.4 représente la distribution d'une loi normale de moyenne (= -0.148) et d'écart-type (= 0.065) estimés par l'approche Bayésienne. A partir de cette distribution, la probabilité a posteriori que l'effet du traitement (Log(HR)) soit inférieur à zéro est de 98.8% (aire grisée Figure 4.4, Tableau 4.3). Cette probabilité est cohérente avec la p-valeur obtenue avec l'analyse fréquentiste ($0.012 \times 2 = 0.024$, proche de la p-valeur 0.023 du test bilatéral).

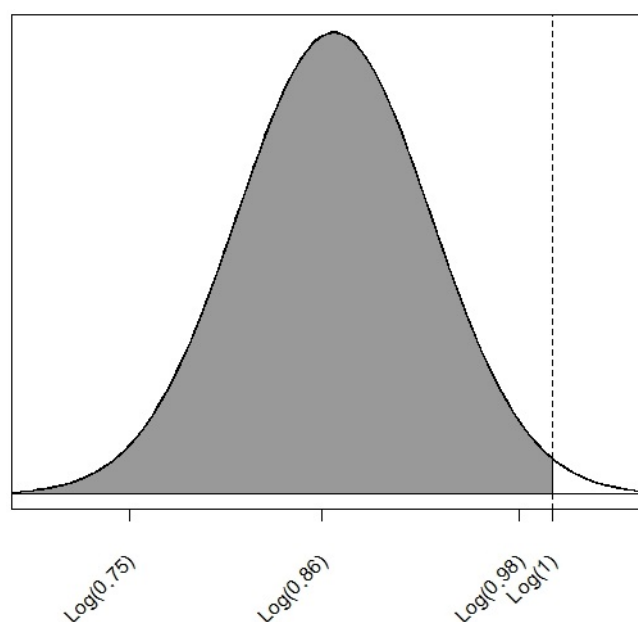


FIGURE 4.4 – Représentation de la probabilité a posteriori de l'effet du traitement $P(\text{Log}(\text{HR}) < 0$), dans l'essai IALT

Probabilité a posteriori de l'effet du traitement (Log(HR)) estimée par l'analyse Bayésienne à partir des données de l'essai IALT (fonction de densité d'une loi normale de moyenne = -0.148 et écart-type = 0.065) et d'une distribution a priori non-informative ($\text{Log}(\text{HR}) \sim N(0,100)$). Dans cet essai $P(\text{HR} < 1) = 98.8\%$, ce qui est cohérent à la p-valeur de 0.023 de l'analyse fréquentiste ($0.012 \times 2 = 0.024$). Sur l'axe des abscisses sont représentés les bornes de l'intervalle de crédibilité à 95% [$\text{Log}(0.75)$; $\text{Log}(0.98)$], la médiane ($\text{Log}(0.86)$), et la borne d'absence d'effet traitement ($\text{Log}(1)$).

4.3. Résultats

Bien que les deux essais aient des p-valeurs proches ($p = 0.026$ pour l'essai EE99-R2Loc et $p = 0.023$ pour l'essai IALT), la taille d'effet de chaque traitement est très différente, comme en témoigne déjà l'analyse fréquentiste avec des estimations de HR très différentes, 0.64 et 0.88 respectivement (Tableau 4.1).

L'analyse Bayésienne permet de caractériser la taille de l'effet traitement par d'autres mesures. Le Tableau 4.3 présente les probabilités a posteriori que le HR soit inférieur à différents seuils, avec notamment le seuil équivalent au rejet de l'hypothèse nulle ($HR < 1$) et les seuils équivalents aux hypothèses alternatives des essais IALT ($HR < 0.85$) et EE99-R2Loc ($HR < 0.60$). Ainsi, dans l'essai IALT, la probabilité a posteriori d'un effet du traitement inférieur à 0.85 n'est pas très importante (41.2%). La principale masse de la distribution se situe entre 1 et 0.85 ($P(1 < HR < 0.85) = 57.6\%$) ce qui montre un effet du traitement très modeste. En revanche, l'essai EE99-R2Loc présente une probabilité a posteriori de 68% pour un $HR < 0.7$, et de 38.4% pour un $HR < 0.6$, ce qui est très encourageant.

Même si l'analyse fréquentiste donne certaines indications sur la taille d'effet, cet exemple démontre que la p-valeur, souvent mal interprétée, ne renseigne en rien le lecteur sur la taille d'effet du traitement et donc sur la probabilité d'un certain bénéfice pour le patient. De plus, l'intervalle de confiance présenté dans une analyse fréquentiste, même s'il semble être égal à l'intervalle de crédibilité lorsqu'une distribution a priori non-informative est utilisée, ne peut être interprété de la même manière sur le plan formel.

Essais	P(HR < 1)	P(HR < 0.85)	P(HR < 0.7)	P(HR < 0.6)
EE99-R2Loc	0.988	0.927	0.684	0.384
IALT	0.988	0.412	< 0.001	< 0.001

TABLEAU 4.3 – Probabilités a posteriori d'avoir un certain bénéfice pour les essais EE99-R2Loc et IALT

Au terme de l'essai HERBY, les investigateurs ont conclu à l'inefficacité du traitement, avec une tendance délétère, comparé au traitement contrôle avec un HR = 1.44 IC95% [0.90-2.30] ($p = 0.132$). L'effet du traitement a posteriori, estimé par l'analyse Bayésienne utilisant une distribution a priori non-informative (Figure 4.5), est égal à 0.364, correspondant à un HR = 1.44, avec une erreur standard $se = 0.241$, permettant d'estimer l'intervalle de crédibilité à 95% suivant ICr95% [-0.105 ; 0.839] pour le Log(HR), équivalent à ICr95% [0.90 ; 2.31].

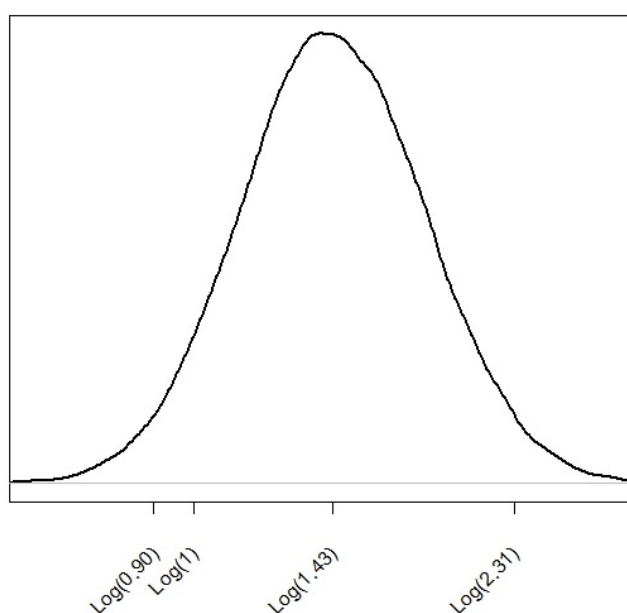


FIGURE 4.5 – Distribution a posteriori empirique du Log(HR) de l'essai HERBY, estimée par un modèle de Cox Bayésien

Distribution a posteriori empirique de l'effet du traitement (Log(HR)) estimée par un modèle de Cox Bayésien au travers d'un processus MCMC, à partir des données de l'essai HERBY et d'une distribution a priori non-informative ($Log(HR) \sim N(0, 100)$). Sur l'axe des abscisses sont représentés les bornes de l'intervalle de crédibilité à 95% [Log(0.90); Log(2.31)], la médiane (Log(1.43)), et la borne d'absence d'effet traitement (Log(1)).

La Figure 4.6 représente la distribution d'une loi normale de moyenne (= 0.364) et d'écart-type (= 0.241) estimés par l'approche Bayésienne.

À partir de cette distribution, la probabilité a posteriori que l'effet du traitement (Log(HR)) soit inférieur à zéro est de 6.5% (aire grisée Figure 4.6). Cette probabilité est cohérente avec

la p-valeur obtenue avec l'analyse fréquentiste $0.065 \times 2 = 0.130$ proche de la p-valeur 0.132 du test bilatéral.

De plus, la probabilité d'un effet délétère tel que le risque d'événement est augmenté de 20% ($HR > 1.2$) est de 77.5%. Cette probabilité reste même assez élevée (54.5%) pour un $HR > 1.4$ (risque d'événement augmenté de 40%).

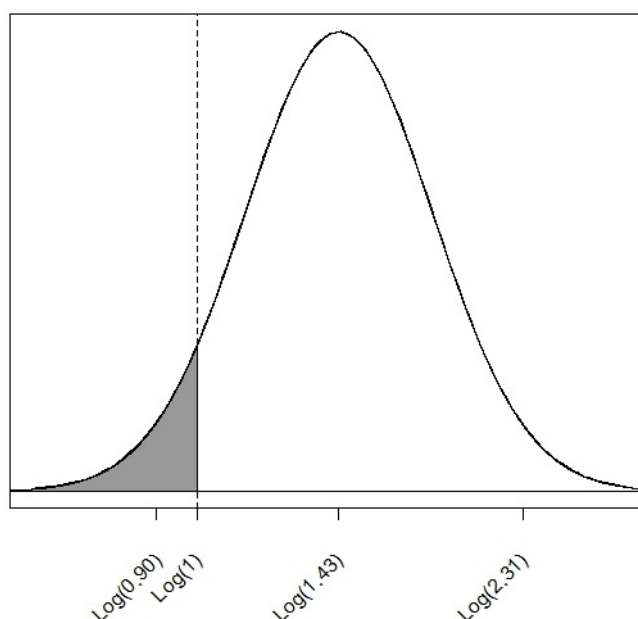


FIGURE 4.6 – Représentation de la probabilité a posteriori de l'effet du traitement $P(\text{Log}(HR) < 0)$, dans l'essai HERBY

Probabilité a posteriori de l'effet du traitement ($\text{Log}(HR)$) estimée par l'analyse Bayésienne à partir des données de l'essai HERBY (fonction de densité d'une loi normale de moyenne = 0.364 et écart-type = 0.241) et d'une distribution a priori non-informative ($\text{Log}(HR) \sim N(0,100)$). Dans cet essai $P(HR < 1) = 0.065$, ce qui est cohérent à la p-valeur de 0.132 de l'analyse fréquentiste ($0.065 \times 2 = 0.130$). Sur l'axe des abscisses sont représentés les bornes de l'intervalle de crédibilité à 95% [$\text{Log}(0.90)$; $\text{Log}(2.31)$], la médiane ($\text{Log}(1.43)$), et la borne d'absence d'effet traitement ($\text{Log}(1)$).

4.3.2 Analyse Bayésienne avec distributions a priori archétypales : analyse de sensibilité

L'analyse de sensibilité, utilisant différentes distributions a priori, nous permet d'évaluer si les résultats de l'essai peuvent être considérés suffisamment persuasifs pour que des cliniciens sceptiques ou enthousiastes soient convaincus.

Nous avons donc utilisé une distribution a priori sceptique et enthousiaste afin d'estimer la distribution a posteriori de l'effet du traitement. Les résultats de l'analyse avec une distribution a priori non informative présentés dans la partie précédente sont repris ici pour mieux illustrer la contribution d'une distribution a priori informative.

Dans le cas de l'essai EE99-R2Loc (Figure 4.7), qui a montré un effet bénéfique du traitement significatif avec l'analyse fréquentiste, l'objectif de cette analyse Bayésienne avec une distribution a priori archétypale est de convaincre les cliniciens sceptiques.

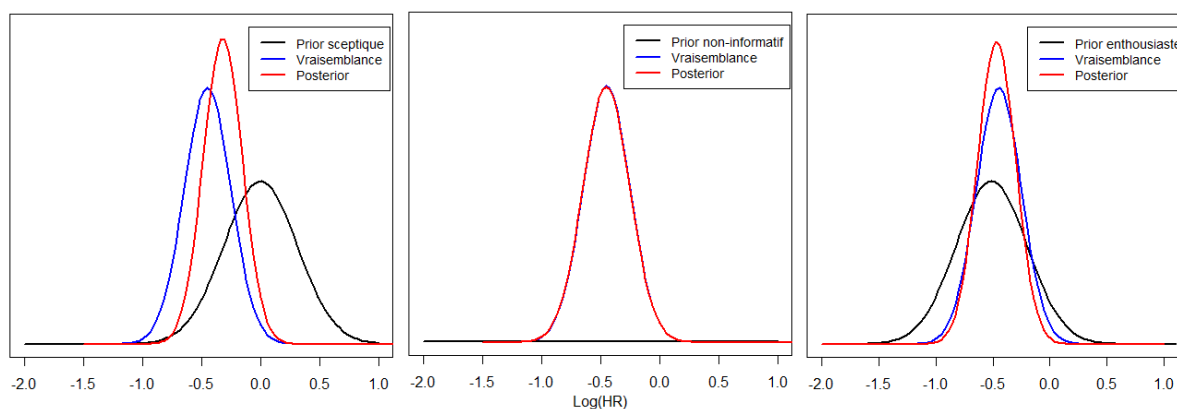


FIGURE 4.7 – Représentation de la distribution a priori, de la vraisemblance, et de la distribution a posteriori de l'effet traitement (Log(HR)) dans l'essai EE99-R2Loc

Le graphique de gauche représente la distribution a priori sceptique, la vraisemblance, et la distribution a posteriori qui en découle. Le graphique du milieu représente une distribution a priori non-informative, la vraisemblance, et la distribution a posteriori associée. Le graphique de droite représente une distribution a priori enthousiaste, la vraisemblance, et la distribution a posteriori qui en découle. Les graduations de l'axe des abscisses correspondent au Log(HR).

4.3. Résultats

La probabilité que le HR soit inférieur à 1, lorsqu'une distribution a priori sceptique est utilisée, est égale à 97.3% (Tableau 4.4). De plus, la probabilité d'un effet du traitement cliniquement significatif (défini ici comme < 0.7) reste élevée (41.1%).

Grâce à l'analyse Bayésienne, nous pouvons voir que les résultats obtenus dans l'essai EE99-R2Loc peuvent être suffisamment persuasifs pour convaincre des cliniciens sceptiques quant à l'effet du traitement sur la survie sans événement.

Dans cette analyse, le poids donné à la distribution a priori peut être exprimé en termes de pseudo-observations (43 pseudo-événements) à comparer aux 105 événements de l'analyse (41%). Ce poids relatif est légèrement supérieur au poids prévu si l'analyse avait été réalisée plus tardivement avec le nombre d'événements initialement prévu (124 événements).

Distributions a priori	P(HR < 1)	P(HR < 0.7)	P(HR < 0.6)
Non-informatif	0.988	0.681	0.379
Sceptique	0.973	0.411	0.124
Enthousiaste	0.997	0.746	0.400

TABLEAU 4.4 – Essai EE99-R2Loc : Probabilités a posteriori d'avoir un certain bénéfice pour différentes valeurs seuils et selon différentes distributions a priori.

Le cas de l'essai IALT est un peu différent, bien que cet essai ait également montré un effet du traitement significatif avec l'analyse fréquentiste. L'objectif est, tout comme l'essai précédent, de convaincre les sceptiques (Figure 4.8).

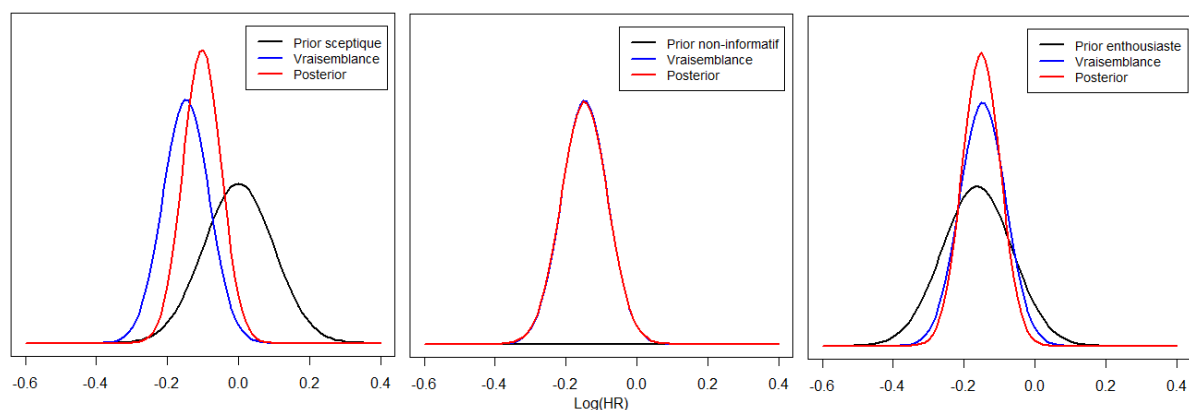


FIGURE 4.8 – Représentation de la distribution a priori, de la vraisemblance, et de la distribution a posteriori de l'effet traitement (Log(HR)) dans l'essai IALT

Le graphique de gauche représente la distribution a priori sceptique, la vraisemblance, et la distribution a posteriori qui en découle. Le graphique du milieu représente une distribution a priori non-informative, la vraisemblance, et la distribution a posteriori associée. Le graphique de droite représente une distribution a priori enthousiaste, la vraisemblance, et la distribution a posteriori qui en découle. Les graduations de l'axe des abscisses correspondent au Log(HR).

Comme le montre le Tableau 4.5, malgré le fait qu'à première vue les sceptiques seraient convaincus qu'un effet du traitement existe ($P(HR < 1) = 97.2\%$), ils ne seraient probablement pas convaincus qu'un effet cliniquement pertinent existe. En effet, la probabilité que le HR soit inférieur à 0.85 n'est que de 13.3%, et inférieur à 0.001% pour un HR inférieur à 0.7.

Distributions a priori	P(HR < 1)	P(HR < 0.85)	P(HR < 0.7)
Non-informatif	0.989	0.456	< 0.001
Sceptique	0.972	0.133	< 0.001
Enthousiaste	0.998	0.420	< 0.001

TABLEAU 4.5 – Essai IALT : Probabilités a posteriori d'avoir un certain bénéfice pour différentes valeurs seuils et selon différentes distributions a priori

Le poids donné à la distribution a priori dans cette analyse peut être exprimé en termes de pseudo-observations (412 pseudo-événements) à comparer aux 973 événements de l'analyse (42%). Ce poids relatif est supérieur au poids prévu si l'analyse avait été réalisée plus tardivement avec le nombre d'événements initialement prévu (1 600 événements).

L'essai HERBY (Figure 4.9) a, lui, échoué à démontrer un effet bénéfique du traitement significatif avec l'analyse fréquentiste, une tendance délétère étant même observée. L'objectif est donc de convaincre les enthousiastes de ce résultat.

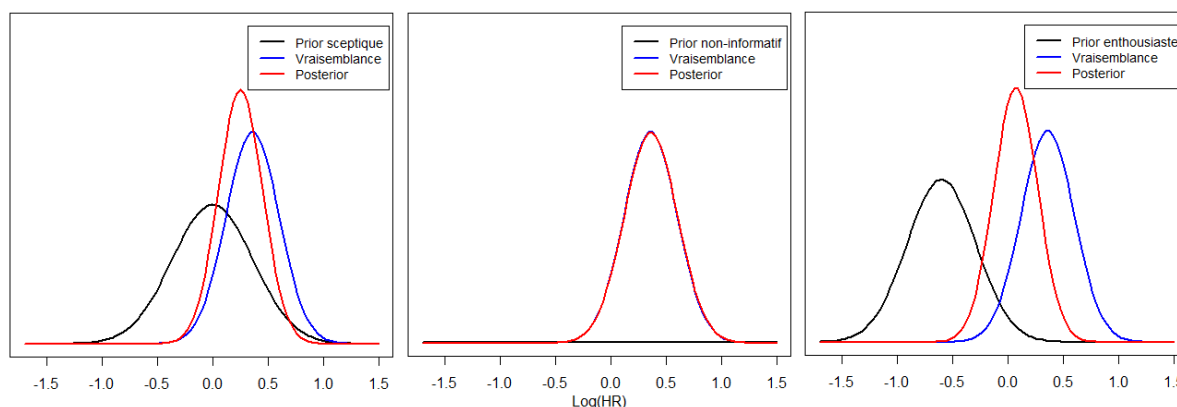


FIGURE 4.9 – Représentation de la distribution a priori, de la vraisemblance, et de la distribution a posteriori de l'effet traitement (Log(HR)) dans l'essai HERBY

Le graphique de gauche représente la distribution a priori sceptique, la vraisemblance, et la distribution a posteriori qui en découle. Le graphique du milieu représente une distribution a priori non-informative, la vraisemblance, et la distribution a posteriori associée. Le graphique de droite représente une distribution a priori enthousiaste, la vraisemblance, et la distribution a posteriori qui en découle. Les graduations de l'axe des abscisses correspondent au Log(HR).

Dans cet essai, la probabilité a posteriori qu'il existe un effet du traitement (probabilité que le HR soit inférieur à 1), lorsqu'une distribution a priori enthousiaste est utilisée, est égale à 35.6% (Tableau 4.6). De plus, la probabilité a posteriori d'un effet du traitement délétère cliniquement inquiétant ($HR > 1.1$) est de plus de 45%.

Grâce à l'analyse Bayésienne, il nous semble que les résultats obtenus dans l'essai HERBY peuvent être suffisamment persuasifs pour convaincre des cliniciens enthousiastes quant à l'inefficacité du traitement, voir à son effet délétère.

À noter que, dans cet essai, la distribution a priori sceptique est en réalité plus "optimiste" que la vraisemblance, puisqu'elle est centrée sur l'absence d'effet (alors que la vraisemblance est centrée sur un HR de 1.44). Dans ce cas de figure, la probabilité a posteriori d'un effet du traitement délétère supérieur à 1.2 est de 64% et s'élève à 34% pour un effet du traitement supérieur à 1.4.

Distributions a priori	P(HR < 1)	P(HR > 1.1)	P(HR > 1.2)	P(HR > 1.4)
Non-informatif	0.065	0.868	0.775	0.545
Sceptique	0.103	0.785	0.638	0.338
Enthousiaste	0.356	0.456	0.292	0.093

TABLEAU 4.6 – Essai HERBY : Probabilités a posteriori d'avoir un certain effet du traitement pour différentes valeurs seuils et selon différentes distributions a priori

Le poids donné à la distribution a priori équivaut à 31 pseudo-événements, à comparer aux 75 événements de l'analyse (41%). Ce poids est légèrement supérieur au poids prévu si l'analyse avait été réalisée plus tardivement avec le nombre d'événements initialement prévu (89 événements).

4.4 Discussion

Nous avons ré-analysé trois essais cliniques, afin d'illustrer la manière dont l'approche Bayésienne peut être utilisée pour enrichir l'interprétation des résultats.

L'utilisation d'une loi a priori non-informative permet de retrouver des résultats cohérents avec ceux de l'analyse fréquentiste, mais surtout de faire une description complète de la distribution a posteriori de l'effet du traitement, et ainsi de donner la probabilité a posteriori que l'effet relatif du traitement soit supérieur, ou inférieur, à une valeur donnée. Les résultats sont ainsi plus facilement interprétables par le clinicien, qui a une vision sur l'ensemble de la distribution du HR. Cette perspective est particulièrement intéressante si l'on cherche à identifier de nouveaux traitements dont l'effet n'est pas simplement "non nul", mais associé à un certain effet clinique. L'analyse Bayésienne nous permet également de décrire la possibilité d'un effet délétère du traitement quand le résultat observé est opposé à ce que l'on espérait, comme c'est le cas pour l'essai HERBY.

Compléter cette analyse par l'utilisation de distributions a priori sceptique et enthousiaste permet de savoir si un clinicien a priori sceptique ou enthousiaste au début de l'essai serait convaincu par les résultats (en supposant que toute sa conviction repose sur l'analyse du HR). Nous avons vu que dans le cas de l'essai EE99-R2Loc, des cliniciens qui auraient été sceptiques au début de l'essai devraient être convaincus de l'efficacité du traitement expérimental, et même d'un effet cliniquement significatif. À l'inverse, dans l'essai IALT, même si des cliniciens sceptiques étaient convaincus d'un effet non nul du traitement, ils ne seraient pas convaincus qu'un bénéfice cliniquement significatif existe. Dans cet essai, même des cliniciens a priori enthousiastes, ne seraient pas convaincus d'un bénéfice cliniquement pertinent. En ce qui concerne l'essai HERBY, des cliniciens enthousiastes n'auraient pas de difficulté à être convaincus de l'absence d'effet du traitement.

Dans ce premier travail, nous avons limité l'analyse à un seul paramètre de variance pour les distributions a priori sceptiques et enthousiastes, tel que recommandé par Freedman et al. [100], basé sur les paramètres de design et non sur le nombre d'événements observés.

Ce choix peut être discuté car son influence sur les résultats, et donc sur la conclusion de l'essai, peut être importante.

Il nous semble indispensable d'estimer le nombre de pseudo-événements (ou pseudo-patients) que représente la variance a priori choisie, et ainsi de le mettre en regard avec le nombre d'événements observés dans l'essai en cours (entre 41 et 42% dans les trois analyses réalisées).

Sur la base d'une analyse fréquentiste, il est hélas fréquent de voir les résultats d'un essai réduit à une conclusion binaire, positif *versus* négatif, à partir de la seule p-valeur en oubliant de discuter la quantité d'effet observé. De plus, la seule estimation du hazard ratio assortie de son intervalle de confiance ne rend pas complètement compte de la distribution possible de l'effet du traitement. Grâce à l'analyse Bayésienne complémentaire, en considérant deux essais avec des p-valeurs assez proches en fréquentiste, l'analyse de l'essai IALT permet de conclure à un effet significatif du traitement, mais relativement faible, sans espoir d'un bénéfice clinique important, tandis-que l'essai EE99-R2Loc est associé à un possible effet majeur du traitement. Il nous semble que l'analyse Bayésienne est plus explicite pour décrire l'effet du traitement que l'on peut espérer au terme d'un essai.

Comme déjà discuté, plusieurs groupes d'experts ont récemment évoqué la nécessité d'estimer l'importance du bénéfice clinique du traitement évalué dans un essai [6–8]. L'approche Bayésienne pourrait constituer un élément de réponse à ce besoin, particulièrement important dans la prise de décision pour les autorités de santé, qui intègre par ailleurs d'autres éléments.

L'approche Bayésienne permettrait également de mieux interpréter les essais classés comme négatifs car non significatifs au seuil habituel de 5%, que ce soit pour conclure à un possible bénéfice mais avec de l'incertitude que l'on quantifie, ou à l'inverse, comme c'est le cas pour l'essai HERBY, à un possible effet délétère. Ainsi, en relisant la conclusion des auteurs de l'essai HERBY, "l'ajout de bevacizumab n'a pas amélioré la survie sans événement des patients", il nous semble que celle-ci pourrait être enrichie par l'approche

Bayésienne. La discussion devrait prendre en compte la probabilité assez élevée d'un effet délétère, ce qui n'est pas mentionné en tant que tel dans la publication princeps de l'article [173].

Parmi les limites de notre travail, nous avons conscience que l'analyse du seul critère de jugement principal est réductrice par rapport à l'ensemble des informations apportées par un essai. D'autres éléments doivent être pris en compte dans l'interprétation d'un essai, tels que l'effet du traitement sur les autres critères d'efficacité et la toxicité, la balance bénéfice risque. Il faut également prendre en compte, la qualité de l'essai évaluée par ses biais éventuels et sa validité externe, plus souvent discutés dans les essais de petite taille que dans les grands essais.

Par ailleurs, nous n'avons considéré que trois essais ce qui limite la généralisabilité de notre conclusion.

Enfin, nous avons prévu une restitution des résultats auprès des cliniciens pour discuter de l'apport de cette analyse Bayésienne et de son interprétation. Ceci reste à faire.

— Chapitre 5 —

Discussion générale

5.1 Synthèse des résultats et apport de la thèse

L'objectif de ce travail de recherche était d'évaluer l'apport de l'approche Bayésienne dans la conduite et l'analyse des essais cliniques avec données censurées, en particulier dans le contexte des maladies rares. Après une revue de la littérature visant à faire un état des lieux de l'implémentation dans la pratique des méthodes Bayésiennes dans des essais cliniques avec un critère censuré, nous avons développé un design Bayésien pour un essai randomisé réel dans une pathologie rare permettant d'intégrer des données historiques de différents types. Enfin, nous avons exploré la manière dont l'analyse Bayésienne peut compléter l'interprétation des résultats d'une analyse fréquentiste d'un essai randomisé avec critère de survie.

Dans la première partie (Chapitre 2), la revue de la littérature a mis en évidence que peu d'essais cliniques ont utilisé les statistiques Bayésiennes pour analyser des données censurées. Contrairement aux nombreux développements méthodologiques en Bayésien, l'utilisation d'une approche Bayésienne pour l'analyse d'un essai clinique en survie reste donc marginale. Ce résultat est cohérent avec les revues de la littérature portant sur l'implémentation de l'approche Bayésienne dans d'autres contextes [145, 146]. Plusieurs raisons peuvent expliquer cette faible utilisation de l'approche Bayésienne. Gönen [40]

suggère que non seulement les cliniciens mais également les statisticiens ne sont pas familiers de l'approche Bayésienne. Si les outils informatiques permettant les analyses Bayésiennes se sont grandement développés dans les dernières années, leur utilisation reste limitée et des difficultés techniques persistent. De plus, la subjectivité inhérente à l'intégration d'une distribution a priori dans un résultat d'étude reste discutable pour beaucoup, que ce soit les statisticiens eux-mêmes, les cliniciens, ou les régulateurs.

Lorsqu'une approche Bayésienne a été utilisée pour l'analyse d'un essai clinique de survie, elle était souvent très peu détaillée. Des recommandations pour le rapport d'analyses Bayésiennes existent (BayesWatch [66] et ROBUST [143]), mais elles ont rarement été suivies dans les articles inclus dans cette revue. Les résultats de cette revue nous amènent à suggérer que le rapport des analyses Bayésiennes dans les essais cliniques pourrait être amélioré, notamment en définissant mieux la distribution a priori utilisée, en clarifiant l'information utilisée pour la formuler, en précisant le modèle de survie utilisé, la méthode d'estimation de ses paramètres, et enfin en explicitant les règles de décision planifiées. Il est possible que les contraintes de nombre de mots imposées par les journaux, aient pu limiter le niveau de détails recommandé; cependant, cette contrainte est factice compte tenu de la possibilité de fournir des informations supplémentaires dans des annexes. Néanmoins, nous pensons qu'une extension du CONSORT (*Consolidated Standards of Reporting Trials*) [175] pour le rapport des essais cliniques Bayésiens pourrait être développée, comprenant des points spécifiques en cas d'analyse de la survie. Une meilleure formation et de plus amples recommandations sur les méthodes Bayésiennes pourraient augmenter la qualité du rapport des analyses de survie Bayésiennes dans les essais cliniques.

Dans la seconde partie (Chapitre 3), nous avons développé un design original Bayésien incorporant des données historiques pour un essai randomisé réel avec un critère censuré, dans le cadre d'une pathologie rare (ostéosarcome). La distribution a priori intégrait des données historiques agrégées sur l'effet relatif du traitement, mais également des données historiques individuelles sur le bras contrôle. À notre connaissance, l'essai développé est le premier essai clinique de survie intégrant des données historiques de ce type. Bien que

l'intérêt de l'utilisation de données historiques ne soit pas nouvelle [78], leur intégration est complexe, en particulier lorsque le critère considéré est un critère censuré [169] et que plusieurs types de données historiques sont à prendre en compte. Dans l'essai Bayésien proposé, le *power prior*, développé par Ibrahim et al. [43], et le *mixture prior*, tel que présenté par Mutsvari et al. [91], ont été utilisés pour incorporer respectivement des informations historiques issues de données individuelles et issues de données agrégées. Nous avons réalisé une étude de simulations pour évaluer les caractéristiques opératoires de ce design et calibrer ses paramètres. De nombreux scénarios ont été considérés pour représenter différents degrés de commensurabilité entre les données historiques et les données de l'essai actuel.

Les résultats des simulations nous ont permis de constater que l'incorporation de données historiques sur le bras contrôle semble avoir peu d'impact sur la puissance de l'essai lorsque les données historiques et actuelles sont commensurables. En revanche, le biais potentiel en cas de non-commensurabilité est important. L'incorporation de ce type de données pour informer l'estimation de l'effet du traitement d'un essai reste donc un choix discutable. Par ailleurs, les gains de puissance et de précision possibles en incorporant des informations historiques commensurables avec les données du nouvel essai, doivent être mis en balance avec le risque de biaiser les estimations, et de perdre en puissance si les données ne sont pas commensurables. Le poids alloué aux données historiques doit être choisi avec attention en fonction de ce compromis au travers d'une étude de simulations. À noter que le choix du paramètre de pondération du *mixture prior* est conditionné par le choix de la variance de la composante vague de cette distribution a priori. Dans la pratique, nous recommandons aux statisticiens planifiant d'utiliser un *mixture prior* pour un nouvel essai clinique Bayésien, de réaliser des simulations afin de calibrer la variance de la composante vague, et son poids $(1-\omega)$, pour trouver les valeurs produisant les caractéristiques opératoires les plus favorables.

Dans la troisième partie (Chapitre 4), nous nous sommes intéressés à l'expression des résultats issus d'une approche Bayésienne, possible même en l'absence de données

historiques pertinentes. Cette approche permet non seulement de faire une description complète de la distribution de l'effet du traitement, à partir de laquelle on peut estimer la probabilité que l'effet du traitement soit inférieur à un seuil défini comme cliniquement pertinent, mais elle permet également d'exprimer ces résultats sous une forme facilement compréhensible par les cliniciens.

Nous avons illustré cette approche à travers la ré-analyse de trois essais randomisés publiés en oncologie : deux essais concluant à un bénéfice significatif du traitement, au seuil alpha 5%, et un essai ne mettant pas en évidence d'effet bénéfique significatif, mais une tendance délétère. L'analyse Bayésienne donne un aperçu du bénéfice possible pour le patient et permet de compléter l'interprétation fréquentiste trop souvent limitée à la p-valeur, classant les résultats en significatifs/non-significatifs.

L'utilisation de distributions a priori sceptique et enthousiaste en complément de la distribution a priori non-informative, permet de représenter la variété d'opinions de la communauté clinique, et ainsi de rendre compte de l'impact du nouvel essai sur les cliniciens a priori sceptiques ou enthousiastes. Cette analyse de sensibilité permet de compléter la conclusion de l'essai en le confrontant à ces opinions "extrêmes".

5.2 Limites

Sans rentrer dans le détail des limites propres à chaque partie de ce travail de thèse, il nous semble important de rappeler certains points de discussion relatifs à l'élaboration ou l'analyse d'un essai Bayésien.

Le design Bayésien que nous avons proposé dans le Chapitre 3, utilise des distributions a priori non dynamiques, ce qui est discutable. En effet, nous avons fait le choix d'utiliser des paramètres de pondération (α_0 et ω) fixes bien que plusieurs auteurs recommandent l'utilisation de distributions a priori dynamiques afin de mieux prendre en compte de potentielles différences entre les données historiques et actuelles [153, 169]. Il est possible que l'utilisation d'un *power prior* avec un paramètre de pondération dynamique, plutôt

que fixe, aurait amélioré le contrôle de l'erreur de type I puisque l'analyse aurait plus vite répondu à un conflit entre les données. Cependant, une version dynamique du *power prior* pourrait mener à une pénalisation plus importante des données historiques contrôles, dans le cadre des essais de petite taille. En effet, dans ce contexte, des différences entre les données contrôles historiques et actuelles, dues à des variations aléatoires plus qu'à de vraies différences entre les paramètres des études, peuvent être observées [176, 177].

Nous avons tenté de rendre l'étude de simulations évaluant le design Bayésien proposé la plus large possible en considérant différents types de scénarios. Cependant, les scénarios que nous avons évalués restaient tout de même proches de l'essai Sarcome-13. Par conséquent, des simulations supplémentaires sont nécessaires afin d'améliorer la généralisation de nos résultats à tous les essais portant sur des pathologies rares, et mesurant un critère censuré. À ce titre, il serait nécessaire d'évaluer l'impact du type de modélisation utilisé en considérant une plus large variété de formes de distributions de survie que celles considérées dans nos jeux de données historiques hypothétiques.

Nous n'avons également considéré que deux types de modèles pour l'analyse du nouvel essai (modèle de Weibull et exponentiel piecewise). D'autres modèles de survie flexibles [161] n'ont pas été évalués car nous avons estimé qu'ils seraient trop complexes à appliquer aux essais de petites tailles.

Nous pensons que le choix de travailler avec les modèles de Weibull et exponentiel piecewise est raisonnable car ils permettent, en faisant varier leurs paramètres, d'ajuster une grande variété de formes de survie différentes.

Lors de la conception d'un essai, il faut tenir compte du temps et des ressources nécessaires pour réaliser de telles simulations. Au-delà des questions de temps de calcul et des infrastructures informatiques, il faut également considérer l'expertise biostatistique nécessaire à la réalisation de ce type d'études de calibration. Il reste du travail avant de disposer de packages, intégrant les différentes options, qui soient accessibles à des biostatisticiens non spécialistes du sujet.

Notre travail sur l'apport de l'approche Bayésienne dans l'expression des résultats s'est limité à la ré-analyse de trois essais cliniques, en considérant les données de l'analyse finale. D'autres exemples seraient nécessaires pour illustrer d'autres situations, telles que celle d'un essai de petite taille avec un résultat non significatif au seuil 5% mais avec une estimation en faveur d'un bénéfice, ou encore les essais de non-infériorité, non couverts par ce travail. Par ailleurs, l'illustration par des exemples ne vaut pas démonstration. Une étude de simulations est nécessaire afin d'évaluer de manière plus précise les règles de décision possibles et les caractéristiques opératoires d'essais qui les implémenteraient.

De plus, nous n'avons pas évalué dans la pratique auprès des différentes parties prenantes (investigateurs impliqués dans les essais ou cliniciens sans pratique de la recherche clinique, promoteurs, patients...) la compréhension réelle des résultats d'un essai lorsque ceux-ci sont exprimés au travers de l'approche Bayésienne. En effet, même si les résultats fournis par cette dernière semblent plus abordables du point de vue du statisticien, les concepts de probabilité et de distribution ne sont pas forcément des notions plus accessibles pour les non initiés.

Une critique possible de notre travail porte sur la réduction de l'analyse d'un essai à l'estimation du hazard ratio d'événement ou de décès. Tout d'abord, il nous semble que la quantification de l'effet d'un traitement en termes de réduction de relative de risques n'est pas simple d'interprétation. D'autres estimateurs de l'effet d'un traitement pourraient être étudiés pour quantifier le bénéfice "absolu", tels que le gain moyen en survie restreinte, ou méthode *restricted mean survival time difference* (rmstd) [161]), mais à notre connaissance l'analyse Bayésienne de ce type de mesure n'a pas encore été développée. Par ailleurs, nous avons conscience que la conclusion d'un essai ne se limite pas à la seule analyse du critère principal d'efficacité. Outre les critères secondaires d'efficacité, l'évaluation des risques est essentielle afin d'évaluer la balance bénéfice-risque, et d'autres éléments participent à la discussion tels que les éventuels biais et problème de validité externe.

5.3 Conclusion et perspectives

Ce travail de recherche a montré que l'approche Bayésienne est peu utilisée en pratique pour l'analyse d'essais cliniques de survie. Dans le cadre des maladies rares, cela peut surprendre lorsque l'on comprend le potentiel apport de cette approche [14, 29, 36–40]. Cependant, sa complexité dissuade probablement bon nombre de cliniciens et statisticiens. En effet, comme nous l'avons vu dans le Chapitre 3, intégrer des données historiques dans le cadre d'une pathologie rare est complexe compte tenu de l'incertitude qui entoure la distribution du risque de base et l'estimation de l'effet du traitement. Le choix du poids accordé aux données historiques est un vrai point de discussion, et l'utilisation d'une distribution a priori dynamique dans le cadre des pathologies rares n'est pas simple et sujet à discussion. Il serait nécessaire d'évaluer l'impact de l'utilisation d'une distribution a priori fixe par rapport à une distribution a priori dynamique dans ce contexte particulier où un important niveau d'incertitude existe.

Bien que potentiellement complexe et sujet à discussion, le principal apport de l'approche Bayésienne repose sur l'intégration possible de données externes. Contrairement aux essais purement fréquentistes dans lesquels les données historiques ne sont utilisées que pour justifier la mise en route de l'essai et le calcul du nombre de sujets nécessaire, l'approche Bayésienne permet de traduire de façon formelle les connaissances extérieures pour en faire une synthèse avec les données de l'essai lui-même. Fondamentalement, l'approche Bayésienne nous semble plus cohérente avec la démarche médicale : chaque jour, le médecin prend des décisions sur une base de probabilités a posteriori intégrant de l'information de sources multiples.

S'il est classique d'évoquer l'apport possible des données externes dans l'analyse Bayésienne d'un essai clinique, celles-ci se limitent le plus souvent à des données historiques, connues avant l'initiation de l'essai. Une perspective possible à ce travail serait de proposer un design Bayésien intégrant également des données de l'essai lui-même, mais sur des critères secondaires. En effet, pour un même résultat sur le critère de jugement

principal, en particulier s'il est à la limite de la significativité dans l'approche classique, la conclusion semble plus solide si elle est corroborée par des résultats cohérents sur les critères secondaires d'efficacité; à l'inverse, l'effet du hasard est souvent mis en avant quand ce résultat "à la limite de la significativité" est observé de façon isolé parmi toutes les analyses réalisées dans l'essai.

Quand nous avons décidé d'intégrer des données historiques pour le groupe contrôle, une des perspectives envisagées était d'évaluer ensuite la possibilité de déséquilibrer les effectifs de l'essai. Pour un nombre total possible de patients dans l'essai, "l'ajout" des pseudo-observations des données historiques aux patients du groupe contrôle permettrait de réduire la taille de ce groupe et donc d'augmenter le nombre de patients dans le groupe expérimental. Cependant, notre étude de simulations a montré un apport limité de l'incorporation de données historiques sur le groupe contrôle, pour des essais avec une randomisation équilibrée. Quoiqu'il soit peu probable que ce résultat soit modifié, il serait cependant intéressant d'évaluer l'apport des données historiques sur le groupe contrôle quand la randomisation est déséquilibrée.

Actuellement des recommandations sont exprimées par les experts pour prendre en compte la magnitude du bénéfice clinique dans la décision au terme d'un essai [7]. Il nous semble qu'une expression Bayésienne des règles de décision proposées serait plus appropriée que la formulation actuelle. Comment expliquer que l'on conclut à un bénéfice cliniquement pertinent si la borne inférieure de l'intervalle de confiance à 95% du hazard ratio est inférieure à 0.7? Un travail entre experts et biostatisticiens serait nécessaire pour mieux formuler une règle de décision basée sur la distribution de la quantité d'effet. Dans cette perspective, un développement à envisager serait d'étendre l'approche Bayésienne ciblée sur l'estimation du bénéfice relatif, exprimée par le hazard ratio, à l'estimation du gain de survie en termes de différence absolue, plus facilement interprétable. Il nous semblerait intéressant d'étudier les possibilités d'estimation Bayésienne de la différence de moyennes de survie restreinte.

Dans nos travaux, nous avons supposé que l'expression Bayésienne des résultats était plus immédiatement interprétable que l'expression classique avec la p-valeur et l'intervalle de confiance. Cependant, il serait intéressant d'évaluer la compréhension des résultats Bayésiens par les personnes amenées à les interpréter.

En conclusion, ce travail nous a permis de mieux appréhender l'apport possible, mais également les difficultés et les limites d'une approche Bayésienne que ce soit dans la construction d'un nouvel essai ou dans l'analyse de résultats pour un critère de survie. Il s'est concrétisé par le développement de l'essai Bayésien Sarcome-13, qui nous l'espérons, encouragera l'implémentation d'autres essais de ce type, particulièrement attendus dans le contexte des pathologies rares.

Bibliographie

1. EUROPEAN MEDICAL AGENCY. Relevant sources for orphan disease prevalence data. 2014. URL : http://www.ema.europa.eu/docs/en_GB/document_library/Other/2012/07/WC500130297.pdf.
2. EUROPEAN COUNCIL. Council Recommendation of 8 June 2009 on an action in the field of rare diseases. 2009. URL : <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2009:151:0007:0010:EN:PDF>.
3. GATTA G, VAN DER ZWAN JM, CASALI PG et al. Rare cancers are not so rare : the rare cancer burden in Europe. Eur J Cancer 2011 ;47 :2493–2511.
4. FISHER RA. Statistical Methods for Research Workers. Edinburgh, United Kingdom : Oliver & Boyd, 1925. 362 pages.
5. BILLINGHAM L, MALOTTKI K et STEVEN N. Research methods to change clinical practice for patients with rare cancers. Lancet Oncol 2016 ;17 :e70–80.
6. STEWART DJ et KURZROCK R. Cancer : The Road to Amiens. J Clin Oncol 2009 ;27 :328–333.
7. CHERNY NI, SULLIVAN R, DAFNI U et al. A standardised, generic, validated approach to stratify the magnitude of clinical benefit that can be anticipated from anti-cancer therapies : the European Society for Medical Oncology Magnitude of Clinical Benefit Scale (ESMO-MCBS). Ann Oncol 2015 ;26 :1547–1573.

8. SCHNIPPER LE, DAVIDSON NE, WOLLINS DS et al. American Society of Clinical Oncology Statement : A Conceptual Framework to Assess the Value of Cancer Treatment Options. *J Clin Oncol* 2015;33 :2563–2577.
9. IDEAL. Final Summary Report : Integrated Design and Analysis of small population group trials (IDeAl). 2017. URL : <https://cordis.europa.eu/docs/results/602/602552/final1-ideal-project-final-report.pdf>.
10. INSPIRE. Final Report Summary - InSPiRe (Innovative methodology for Small Populations Research). 2017. URL : https://cordis.europa.eu/result/rcn/204338_en.html.
11. ASERIX. Final Summary Report : Advances in Small Trials dEsign for Regulatory Innovation and eXcellence (ASTERIX). 2017. URL : <https://cordis.europa.eu/docs/results/603/603160/final1-final-summary-report-29nov2017.pdf>.
12. HILGERS RD, ROES K, STALLARD N et al. Directions for new developments on statistical design and analysis of small population group trials. *Orphanet Journal of Rare Diseases* 2016;11 :78.
13. FOOD AND DRUG ADMINISTRATION. How FDA Plans to Help Consumers Capitalize on Advances in Science. 2017. URL : <https://blogs.fda.gov/fdavoices/index.php/tag/in-silico-tools/>.
14. CASALI PG, BRUZZI P, BOGAERTS J et al. Rare Cancers Europe (RCE) methodological recommendations for clinical studies in rare cancers : a European consensus position paper. *Ann Oncol* 2015;26 :300–306.
15. BAJARD A, CHABAUD S, CORNU C et al. An in silico approach helped to identify the best experimental design, population, and outcome for future randomized clinical trials. *J Clin Epidemiol* 2016;69 :125–36.
16. PAPPALARDO F, RUSSO G, MUSUAMBA TSHINANU F et al. In silico clinical trials : concepts and early adoptions. *Briefings in Bioinformatics* 2018. (in press).

17. BAEY C et LE DELEY MC. Effect of misspecification of response rates on type I and II errors, in a Phase II Simon design. *Eur J Cancer* 2011 ;47 :1647–52.
18. RATAIN MJ et SARGENT DJ. Optimising the design of phase II oncology trials : The importance of randomisation. *Eur J Cancer* 2009 ;45 :275–80.
19. CHOW SC et CHANG M. *Adaptive Design Methods in Clinical Trials*. Chapman & Hall, 2011. 374 pages.
20. GEHAN EA. The determination of the number of patients required in a preliminary and a follow-up trial of a new chemotherapeutic agent. *J Chronic Dis* 1961 ;13 :346–53.
21. FLEMING TR. One-Sample Multiple Testing Procedure for Pphase II Clinical Trials. *Biometrics* 1982 ;38 :143–51.
22. SIMON R. Optimal Two-Stage Designs for Phase II Clinical Trials. *Control Clin Trials* 1989 ;10 :1–10.
23. BRYANT J et DAY R. Incorporating Toxicity Considerations into the Design of Two-Stage Phase II Clinical Trials. *Biometrics* 1995 ;51 :1372–1383.
24. MANDREKAR SJ et SARGENT DJ. Pick the Winner Designs in Phase II Cancer Clinical Trials. *J Thorac Oncol* 2006 ;1 :5–6.
25. KATZ N. Enriched Enrollment Randomized Withdrawal Trial Designs of Znalgesics : Focus on Methodology. *Clin J Pain* 2009 ;25 :797–807.
26. SIMON N et SIMON R. Adaptive enrichment designs for clinical trials. *Biostatistics* 2013 ;14 :613–625.
27. JOHNSON J et R. T. Food and Drug Administration requirements for approval of new anticancer drugs. *Cancer Treat Rep* 1985 ;69 :1155–9.
28. SAAD E et BUYSE M. Statistical controversies in clinical research : end points other than overall survival are vital for regulatory approval of anticancer agents. *Ann Oncol* 2016 ;27 :373–8.

29. EUROPEAN MEDICAL AGENCY. Guideline on Clinical trials in Small Populations. 2006.
URL : http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003615.pdf.
30. R. S, JOHNSON J, JUSTICE R et al. Review of oncology and hematology drug product approvals at the US Food and Drug Administration between July 2005 and December 2007. *J Natl Cancer Inst* 2010;102 :230–43.
31. CHERNY NI, DAFNI U, BOGAERTS J et al. ESMO-Magnitude of Clinical Benefit Scale version 1.1. *Ann Oncol* 2017;28 :2340–2366.
32. BOGAERTS J, SYDES MR, KEAT N et al. Clinical trial designs for rare diseases : Studies developed and discussed by the International Rare Cancers Initiative. *Eur J Cancer* 2015;51 :271–81.
33. SPOSTO R et STRAM DO. A strategic view of randomized trial design in low-incidence paediatric cancer. *Stat Med* 1999;18 :1183–1197.
34. LE DELEY MC, BALLMAN KV, MARANDET J et al. Taking the long view : How to design a series of Phase III trials to maximize cumulative therapeutic benefit. *Clin Trials* 2012;9 :283–292.
35. BAYAR MA, LE TEUFF G, MICHIELS S et al. New insights into the evaluation of randomized controlled trials for rare diseases over a long-term research horizon : a simulation study. *Stat Med* 2016;35 :3245–58.
36. LILFORD RJ, THORNTON JG et BRAUNHOLTZ D. Clinical trials and rare diseases : a way out of a conundrum. *BMJ* 1995;311 :1621–1625.
37. TAN SB, DEAR KBG, BRUZZI P et al. Strategy for randomised clinical trials in rare cancers. *BMJ* 2003;327 :47–9.
38. BILLINGHAM L, MALOTTKI K et STEVEN N. Small sample sizes in clinical trials : a statistician's perspective. *Clin Invest* 2012;2 :655–657.
39. TUDUR SMITH C, WILLIAMSON PR et BERESFORD MW. Methodology of clinical trials for rare diseases. *Best Pract Res Clin Rheumatol* 2014;28 :247–62.

40. GÖNEN M. Bayesian clinical trials : no more excuses. *Clin Trials* 2009 ;6 :203–204.
41. NIKOLAKOPOULOS S, VAN DER TWEEL I et ROES KC. Dynamic Borrowing through Empirical Power Priors That Control Type I Error. *Biometrics* 2017. (in press).
42. GALWEY NW. Supplementation of a clinical trial by historical control data : is the prospect of dynamic borrowing an illusion ? *Stat Med* 2017 ;36 :899–916.
43. IBRAHIM JG et CHEN MH. Power prior distributions for regression models. *Statist Sci* 2000 ;15 :46–60.
44. THALL PF, URSINO M, BAUDOIN V et al. Bayesian Treatment Comparison Using Parametric Mixture Priors Computed from Elicited Histograms. *Stat Methods Med Res* 2017. (in press).
45. BISWAS S, LIU DD, LEE JJ et al. Bayesian clinical trials at the University of Texas M. D. Anderson Cancer Center. *Clin Trials* 2009 ;6 :205–16.
46. BAYES T. An Essay towards solving a Problem in the Doctrine of Chances. *Philosoph Trans Roy Soc London* 1763 ;53 :370–418.
47. TSIATIS AA. The asymptotic Joint distribution of the efficient scores test for the proportional hazards model calculated over time. *Biometrika* 1981 ;68 :311–5.
48. GILKS W, RICHARDSON S et SPIEGELHALTER D. *Markov Chain Monte Carlo in Practice*. Chapman & Hall : London, 1996. 504 pages.
49. GEMAN S et GEMAN D. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1984 ;6 :721–741.
50. GELFAND AE et SMITH AF. Sampling-Based Approaches to Calculating Marginal Densities. *J Am Stat Assoc* 1990 ;85 :398–409.
51. METROPOLIS N, ROSENBLUTH A, ROSENBLUTH M et al. Equations of State Calculations by Fast Computing Machines. *J Chem Phys* 1953 ;21 :1087–1092.
52. HASTINGS WK. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 1970 ;57 :97–109.

53. GEWEKE J. Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments. In : *Bayesian Statistics 4*. University Press, 1992 :169–193.
54. HEIDELBERGER P et WELCH PD. Simulation run length control in the presence of an initial transient. *Op Research* 1983 ;31 :1109–1144.
55. BERNARDO JM, BERGER JO, DAWID AP et al. *Bayesian Statistics 4*. Oxford University Press : Oxford, 1991. 876 pages.
56. GELMAN A et RUBIN DB. Inference from Iterative Simulation Using Multiple Sequences. *Statist Sci* 1992 ;7 :457–511.
57. BROOKS SP et GELMAN A. General Methods for Monitoring Convergence of Iterative Simulation. *J Comput Graph Statist* 1998 ;7 :434–455.
58. BERRY DA. Bayesian clinical trials. *Nat Rev Drug Discov* 2006 ;5 :27–36.
59. LEE JJ et CHU CT. Bayesian clinical trials in action. *Stat Med* 2012 ;31 :2955–72.
60. WESTOVER MB, WESTOVER KD et BIANCHI MT. Significance testing as perverse probabilistic reasoning. *BMC Med* 2011 ;9 :20.
61. GSPONER T, GERBER F, BORNKAMP B et al. A practical guide to Bayesian group sequential designs. *Pharm Stat* 2014 ;13 :71–80.
62. SAVILLE BR, CONNOR JT, AYERS GD et al. The utility of Bayesian predictive probabilities for interim monitoring of clinical trials. *Clin Trials* 2014 ;11 :485–493.
63. THOMAS A, SPIEGELHALTER D et GILKS W. BUGS : a program to perform Bayesian inference using Gibbs sampling. In : *Bayesian Statistics 4*. Sous la dir. de BERNARDO JM, DAWID AP et SMITH AFM. Oxford University Press : Oxford, 1992 :837–842.
64. PARMAR MKB, SPIEGELHALTER DJ et FREEDMAN LS. The CHART trials : Bayesian design and monitoring in practice. *Stat Med* 1994 ;13 :1297–1312.
65. PARMAR MKB, UNGERLEIDER RS et SIMON R. Assessing Whether to Perform a Confirmatory Randomized Clinical Trial. *J Natl Cancer Inst* 1996 ;88 :1645–51.
66. SPIEGELHALTER DJ, MYLES JP, JONES DR et al. Bayesian methods in health technology assessment : a review. *Health Technol Assess* 2000 ;4 :1–130.

67. LESAFFRE E et LAWSON AB. *Bayesian Biostatistics*. Wiley, 2012. 534 pages.
68. CHALONER K et RHAME FS. Quantifying and documenting prior beliefs in clinical trials. *Stat Med* 2001 ;20 :581–600.
69. SPIEGELHALTER DJ, ABRAMS KR et MYLES JP. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. Wiley, 2004. 408 pages.
70. O’HAGAN A, BUCK CE, DANESHKHAH A et al. *Uncertain Judgements : Eliciting Experts’ Probabilities*. Wiley, 2006. 321 pages.
71. HIANCE A, CHEVRET S et LÉVY V. A practical approach for eliciting expert prior beliefs about cancer survival in phase III randomized trial. *J Clin Epidemiol* 2009 ;62 :431–437.e2.
72. JOHNSON SR, TOMLINSON GA, HAWKER GA et al. A valid and reliable belief elicitation method for Bayesian priors. *J Clin Epidemiol* 2010 ;63 :370–383.
73. ALBERT I, DONNET S, GUIHENNEUC-JOUYAUX C et al. Combining Expert Opinions in Prior Elicitation. *Bayesian Anal* 2012 ;7 :503–532.
74. KINNERSLEY N et DAY S. Structured approach to the elicitation of expert beliefs for a Bayesian-designed clinical trial : a case study. *Pharm Stat* 2013 ;12 :104–13.
75. MOATTI M, ZOHAR S, FACON Y et al. Modeling of experts divergent prior beliefs for a sequential phase III clinical trial. *Clin Trials* 2013 ;10 :505–514.
76. PIBOULEAU L et CHEVRET S. An internet-based method to elicit experts’ beliefs for Bayesian priors : a case study in intracranial stent evaluation. *Int J Technol Assess Health Care* 2014 ;30 :446–453.
77. HAMPSON LV, WHITEHEAD J, ELEFThERIOU D et al. Elicitation of Expert Prior Opinion : Application to the MYPAN Trial in Childhood Polyarteritis Nodosa. *Plos One* 2015 ;10 :e0120981.
78. POCOCK SJ. The combination of randomized and historical controls in clinical trials. *J Chronic Dis* 1976 ;29 :175–88.

79. FOOD AND DRUG ADMINISTRATION. Guidance for industry and FDA staff : Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials. 2010. URL : <http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm071121.pdf>.
80. WADSWORTH I, HAMPSON L et JAKI T. Extrapolation of efficacy and other data to support the development of new medicines for children : A systematic review of methods. *Stat Methods Med Res* 2018 ;27 :398–413.
81. IBRAHIM JG, CHEN MH et CHU H. Bayesian methods in clinical trials : a Bayesian analysis of ECOG trials E1684 and E1690. *BMC Med Res Methodol* 2012 ;12 :183.
82. NEUENSCHWANDER B, BRANSON M et SPIEGELHALTER DJ. A note on the power prior. *Stat Med* 2009 ;28 :3562–3566.
83. HOBBS BP, CARLIN BP, MANDREKAR S et al. Hierarchical Commensurate and Power Prior Models for Adaptive Incorporation of Historical Information in Clinical Trials. *Biometrics* 2011 ;67 :1047–56.
84. HOBBS BP, SARGENT DJ et CARLIN BP. Commensurate Priors for Incorporating Historical Information in Clinical Trials Using General and Generalized Linear Models. *Bayesian Anal* 2012 ;7 :639–674.
85. MURRAY TA, HOBBS BP, LYSTIG TC et al. Semiparametric Bayesian commensurate survival model for post-market medical device surveillance with non-exchangeable historical data. *Biometrics* 2014 ;70 :185–91.
86. KAIZER AM, KOOPMEINERS JS et HOBBS BP. Bayesian Hierarchical Modeling based on Multisource Exchangeability. *Biostatistics* 2018 ;19 :169–184.
87. CAMPBELL G. Bayesian methods in clinical trials with applications to medical devices. *Communications for Statistical Applications and Methods* 2017 ;24 :561–581.
88. NEUENSCHWANDER B, CAPKUN-NIGGLI G, BRANSON M et al. Summarizing historical information on controls in clinical trials. *Clin Trials* 2010 ;7 :5–18.

89. GSTEIGER S, NEUENSCHWANDER B, MERCIER F et al. Using historical control information for the design and analysis of clinical trials with overdispersed count data. *Stat Med* 2013 ;32 :3609–22.
90. SCHMIDLI H, GSTEIGER S, ROYCHOUDHURY S et al. Robust meta-analytic-predictive priors in clinical trials with historical control information. *Biometrics* 2014 ;70 :1023–32.
91. MUTSVARI T, TYTGAT D et WALLEY R. Addressing potential prior-data conflict when using informative priors in proof-of-concept studies. *Pharm Stat* 2016 ;15 :28–36.
92. GELMAN A. Objections to Bayesian statistics. *Bayesian Anal* 2008 ;3 :445–449.
93. MOYÉ LA. Bayesians in clinical trials : Asleep at the switch. *Stat Med* 2008 ;27 :469–82.
94. WASSERSTEIN RL et LAZAR NA. The ASA's Statement on p-Values : Context, Process, and Purpose. *Am Statist* 2016 ;70 :129–133.
95. HUGHES MD. Reporting Bayesian analyses of clinical trials. *Stat Med* 1993 ;12 :1651–1663.
96. GOLIGHER E, TOMLINSON G, HAJAGE D 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 Clinical Trial. *JAMA* 2018. Sous presse.
97. BERRY D. Interim analyses in clinical trials : Classical vs. Bayesian approaches. *Stat Med* 1985 ;4 :521–6.
98. CARLIN BP, CHALONER K, CHURCH T et al. Bayesian approaches for monitoring clinical trials with an application to toxoplasmic encephalitis prophylaxis. *The Statistician* 1993 ;42 :355–367.
99. GEORGE SL, LI C, BERRY DA et al. Stopping a clinical trial early : Frequentist and Bayesian approaches applied to a CALGB trial in non-small-cell lung cancer. *Stat Med* 1994 ;13 :1313–27.

100. FREEDMAN LS, SPIEGELHALTER DJ et PARMAR MK. The What, Why and How of Bayesian clinical trials monitoring. *Stat Med* 1994 ;13 :1371–1383.
101. HEITJAN DF. Bayesian interim analysis of phase II cancer clinical trials. *Stat Med* 1997 ;16 :1791–802.
102. ROSNER GL. Bayesian Monitoring of Clinical Trials with Failure-Time Endpoints. *Biometrics* 2005 ;61 :239–245.
103. JOHNSON VE et COOK JD. Bayesian Design of Single-arm Phase II Clinical Trials with Continuous Monitoring. *Clin Trials* 2009 ;6 :217–26.
104. FAYERS PM, ASHBY D et PARMAR MKB. Tutorial in biostatistics Bayesian data monitoring in clinical trials. *Stat Med* 1997 ;16 :1413–30.
105. LUNN D, JACKSON C, BEST N et al. *The BUGS Book : A Practical Introduction to Bayesian Analysis*. Chapman & Hall/CRC Press. 2012. 399 pages.
106. EUROPEAN MEDICAL AGENCY. Concept paper on extrapolation of efficacy and safety in medicine development. 2013. URL : <http://www.sftox.com/actualites/WC500142358.pdf>.
107. VIELE K, BERRY S, NEUENSCHWANDER B et al. Use of historical control data for assessing treatment effects in clinical trials. *Pharm Stat* 2014 ;13 :41–54.
108. HOGARTH RM. Cognitive Processes and the Assessment of Subjective Probability Distributions. *J Am Statist Assoc* 1975 ;70 :271–289.
109. BRARD C, LE TEUFF G, LE DELEY MC et al. Bayesian survival analysis in clinical trials : What methods are used in practice ? *Clin Trials* 2017 ;14 :78–87.
110. IBRAHIM JG, CHEN MH et SINHA D. *Bayesian Survival Analysis*. New-York : Springer, 2001. 480 pages.
111. ANDERSON H, LUND B, BACH F et al. Single-Agent Activity of Weekly Gemcitabine in Advanced Non-Small-Cell Lung Cancer : A Phase II Study. *J Clin Oncol* 1994 ;12 :1821–1826.

112. LEVITT SH, AEPPLI DM et NIERENGARTEN ME. The Impact of Radiation on Early Breast Carcinoma Survival : A Bayesian analysis. *Cancer* 1996 ;78 :1035–42.
113. MÉNARD S, VALAGUSSA P, PILOTTI S et al. Response to Cyclophosphamide, Methotrexate, and Fluorouracil in Lymph Node–Positive Breast Cancer According to HER2 Overexpression and Other Tumor Biologic Variables. *J Clin Oncol* 2001 ;19 :329–35.
114. ESTEY EH, THALL PF, GILES FJ et al. Gemtuzumab ozogamicin with or without interleukin 11 in patients 65 years of age or older with untreated acute myeloid leukemia and high-risk myelodysplastic syndrome : comparison with idarubicin plus continuous-infusion, high-dose cytosine arabinoside. *Blood* 2002 ;99 :4343–9.
115. GENNARI A, AMADORI D, DE LENA M et al. Lack of Benefit of Maintenance Paclitaxel in First-Line Chemotherapy in Metastatic Breast Cancer. *J Clin Oncol* 2006 ;24 :3912–8.
116. MAKI RG, WATHEN JK, PATEL SR et al. Randomized Phase II Study of Gemcitabine and Docetaxel Compared with Gemcitabine Alone in Patients With Metastatic Soft Tissue Sarcomas : Results of Sarcoma Alliance for Research Through Collaboration Study 002. *J Clin Oncol* 2007 ;25 :2755–2763.
117. CHAMBERLAIN MC et GLANTZ MJ. Interferon-alpha for recurrent World Health Organization grade 1 intracranial meningiomas. *Cancer* 2008 ;113 :2146–51.
118. MIKSAD RA, GÖNEN M, LYNCH TJ et al. Interpreting Trial Results in Light of Conflicting Evidence : A Bayesian Analysis of Adjuvant Chemotherapy for Non–Small-Cell Lung Cancer. *J Clin Oncol* 2009 ;27 :2245–52.
119. MUSS HB, BERRY DA, CIRRINCIONE CT et al. Adjuvant Chemotherapy in Older Women with Early-Stage Breast Cancer. *N Engl J Med* 2009 ;360 :2055–65.
120. NORMAN PH, THALL PF, PURUGGANAN RV et al. A Possible Association Between Aprotinin and Improved Survival After Radical Surgery for Mesothelioma. *Cancer* 2009 ;115 :833–41.

121. PECORELLI S, FAVALLI G, GADDUCCI A et al. Phase III Trial of Observation Versus Six Courses of Paclitaxel in Patients With Advanced Epithelial Ovarian Cancer in Complete Response After Six Courses of Paclitaxel/Platinum-Based Chemotherapy : Final Results of the After-6 Protocol 1. *J Clin Oncol* 2009 ;27 :4642–4648.
122. DE LIMA M, GIRALT S, THALL PF et al. Maintenance therapy with low-dose azacitidine after allogeneic hematopoietic stem cell transplantation for recurrent acute myelogenous leukemia or myelodysplastic syndrome : a dose and schedule finding study. *Cancer* 2010 ;116 :5420–5431.
123. LU C, LEE JJ, KOMAKI R et al. Chemoradiotherapy with or without AE-941 in stage III non-small cell lung cancer : a randomized phase III trial. *J Natl Cancer Inst* 2010 ;102 :859–65.
124. WILBER DJ, PAPPONE C, NEUZIL P et al. Comparison of antiarrhythmic drug therapy and radiofrequency catheter ablation in patients with paroxysmal atrial fibrillation : a randomized controlled trial. *JAMA* 2010 ;303 :333–40.
125. GARCIA-MANERO G, TAMBARO FP, BEKELE NB et al. Phase II trial of vorinostat with idarubicin and cytarabine for patients with newly diagnosed acute myelogenous leukemia or myelodysplastic syndrome. *J Clin Oncol* 2012 ;30 :2204–10.
126. SHARMA M, KHAN H, THALL PF et al. A Randomized Phase II Trial of a Preparative Regimen of Bortezomib, High Dose Melphalan, Arsenic Trioxide and Ascorbic Acid. *Cancer* 2012 ;118 :2507–15.
127. SHULMAN LN, CIRRINCIONE CT, BERRY DA et al. Six cycles of doxorubicin and cyclophosphamide or Paclitaxel are not superior to four cycles as adjuvant chemotherapy for breast cancer in women with zero to three positive axillary nodes : Cancer and Leukemia Group B 40101. *J Clin Oncol* 2012 ;30 :4071–6.
128. CURTIS AB, WORLEY SJ, ADAMSON PB et al. Biventricular pacing for atrioventricular block and systolic dysfunction. *N Engl J Med* 2013 ;368 :1585–93.

129. HENSLEY ML, WATHEN JK, MAKI RG et al. Adjuvant therapy for high-grade, uterus-limited leiomyosarcoma : results of a phase 2 trial (SARC 005). *Cancer* 2013;119 :1555–61.
130. VON MINCKWITZ G, MÖBUS V, SCHNEEWEISS A et al. German adjuvant intergroup node-positive study : a phase III trial to compare oral ibandronate versus observation in patients with high-risk early breast cancer. *J Clin Oncol* 2013 ;31 :3531–9.
131. MORRIS RK, MALIN GL, QUINLAN-JONES E et al. Percutaneous vesicoamniotic shunting versus conservative management for fetal lower urinary tract obstruction (PLUTO) : a randomised trial. *Lancet* 2013 ;382 :1496–1506.
132. UENO H, IOKA T, IKEDA M et al. Randomized phase III study of gemcitabine plus S-1, S-1 alone, or gemcitabine alone in patients with locally advanced and metastatic pancreatic cancer in Japan and Taiwan : GEST study. *J Clin Oncol* 2013 ;31 :1640–8.
133. MOTZER RJ, BARRIOS CH, KIM TM et al. Phase II randomized trial comparing sequential first-line everolimus and second-line sunitinib versus first-line sunitinib and second-line everolimus in patients with metastatic renal cell carcinoma. *J Clin Oncol* 2014 ;32 :2765–2772.
134. KANAKRY CG, O'DONNELL PV, FURLONG T et al. Multi-institutional study of post-transplantation cyclophosphamide as single-agent graft-versus-host disease prophylaxis after allogeneic bone marrow transplantation using myeloablative busulfan and fludarabine conditioning. *J Clin Oncol* 2014 ;32 :3497–505.
135. LE DELEY MC, PAULUSSEN M, LEWIS I et al. Cyclophosphamide compared with ifosfamide in consolidation treatment of standard-risk Ewing sarcoma : results of the randomized noninferiority Euro-EWING99-R1 trial. *J Clin Oncol* 2014 ;32 :2440–8.
136. SATOUCHI M, KOTANI Y, SHIBATA T et al. Phase III study comparing amrubicin plus cisplatin with irinotecan plus cisplatin in the treatment of extensive-disease small-cell lung cancer : JCOG 0509. *J Clin Oncol* 2014 ;32 :1262–8.

137. SHULMAN LN, BERRY DA, CIRRINCIONE CT et al. Comparison of doxorubicin and cyclophosphamide versus single-agent paclitaxel as adjuvant therapy for breast cancer in women with 0 to 3 positive axillary nodes : CALGB 40101 (Alliance). *J Clin Oncol* 2014 ;32 :2311–2317.
138. STEPHENSON JJ, NEMUNAITIS J, JOY AA et al. Randomized phase 2 study of the cyclin-dependent kinase inhibitor dinaciclib (MK-7965) versus erlotinib in patients with non-small cell lung cancer. *Lung Cancer* 2014 ;83 :219–23.
139. GELMAN A, CARLIN JB, STERN HS et al. *Bayesian Data Analysis*. Thrid Edition. Chapman & Hall/CRC press, 2013. 675 pages.
140. THALL PF et SUNG HG. Some extensions and applications of a Bayesian strategy for monitoring multiple outcomes in clinical trials. *Stat Med* 1998 ;17 :1563–80.
141. BROPHY JM et JOSEPH L. Placing trials in context using Bayesian analysis. GUSTO revisited by Reverend Bayes. *JAMA* 1995 ;273 :871–5.
142. SCHWARZ G. Estimating the Dimension of a Model. *Ann Stat* 1978 ;6 :461–464.
143. SUNG L, HAYDEN J, GREENBERG ML et al. Seven items were identified for inclusion when reporting a Bayesian analysis of a clinical study. *J Clin Epidemiol* 2005 ;58 :261–8.
144. ASHBY D. Bayesian statistics in medicine : a 25 year review. *Stat Med* 2006 ;25 :3589–631.
145. PIBOULEAU L et CHEVRET S. Bayesian statistical method was underused despite its advantages in the assessment of implantable medical devices. *J Clin Epidemiol* 2011 ;64 :270–9.
146. CHEVRET S. Bayesian adaptive clinical trials : a dream for statisticians only? *Stat Med* 2012 ;31 :1002–13.

147. SAUNDERS M, DISCHE S, BARRETT A et al. Continuous, hyperfractionated, accelerated radiotherapy (CHART) versus conventional radiotherapy in non-small cell lung cancer : mature data from the randomised multicentre trial. *Radiother Oncol* 1999;52 :137–48.
148. BERRY DA. Bayesian methods in phase III trials. *Drug information journal* 1991 ;25 :345–368.
149. BERRY DA. Introduction to Bayesian methods III : use and interpretation of Bayesian tools in design and analysis. *Clin Trials* 2005;2 :295–300, discussion 301–304, 364–378.
150. GREENHOUSE JB et SELTMAN H. Using prior distributions to synthesize historical evidence : comments on the Goodman-Sladky case study of IVIg in Guillain-Barré syndrome. *Clin Trials* 2005;2 :311–8.
151. PARMAR MKB, SYDES MR et MORRIS TP. How do you design randomised trials for smaller populations ? A framework. *BMC Med* 2016 ;14 :183.
152. LI JX, CHEN WC et SCOTT JA. Addressing prior-data conflict with empirical meta-analytic-predictive priors in clinical studies with historical information. *J Biopharm Stat* 2016 ;26 :1056–1066.
153. GRAVESTOCK I, HELD L et COMBACTE-NET CONSORTIUM. Adaptive power priors with empirical Bayes for clinical trials. *Pharm Stat* 2017 ;16 :349–360.
154. VAIL D, MAC EWEN E, KURZMAN I et al. Liposome-encapsulated muramyl tripeptide phosphatidylethanolamine adjuvant immunotherapy for splenic hemangiosarcoma in the dog : a randomized multi-institutional clinical trial. *Clin Cancer Res* 1995;1 :1165–70.
155. KURZMAN ID, MAC EWEN EG, ROSENTHAL RC et al. Adjuvant therapy for osteosarcoma in dogs : results of randomized clinical trials using combined liposome-encapsulated muramyl tripeptide and cisplatin. *Clin Cancer Res* 1995;1 :1595–601.

156. MEYERS PA, SCHWARTZ CL, KRAILO M et al. Osteosarcoma : A Randomized, Prospective Trial of the Addition of Ifosfamide and/or Muramyl Tripeptide to Cisplatin, Doxorubicin, and High-Dose Methotrexate. *J Clin Oncol* 2005 ;23 :2004–11.
157. MEYERS PA, SCHWARTZ CL, KRAILO MD et al. Osteosarcoma : The Addition of Muramyl Tripeptide to Chemotherapy Improves Overall Survival - A Report From the Children's Oncology Group. *J Clin Oncol* 2008 ;26 :633–8.
158. CHOU AJ, KLEINERMAN ES, KRAILO MD et al. Addition of Muramyl Tripeptide to Chemotherapy for Patients with Newly Diagnosed Metastatic Osteosarcoma : a Report from the Children's Oncology Group. *Cancer* 2009 ;115 :5339–48.
159. KAGER L, PÖTSCHGER U et BIELACK S. Review of mifamurtide in the treatment of patients with osteosarcoma. *Ther Clin Risk Manag* 2010 ;6 :279–286.
160. PIPERNO-NEUMANN S, LE DELEY MC, RÉDINI F et al. Zoledronate in combination with chemotherapy and surgery to treat osteosarcoma (OS2006) : a randomised, multicentre, open-label, phase 3 trial. *Lancet Oncol* 2016 ;17 :1070–1080.
161. ROYSTON P et PARMAR MKB. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Stat Med* 2002 ;21 :2175–97.
162. AKAIKE H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 1974 ;19 :716–723.
163. SCHOENFELD DA. Sample-Size Formula for the Proportional-Hazards Regression Model. *Biometrics* 1983 ;39 :499–503.
164. R CORE TEAM, R FOUNDATION FOR STATISTICAL COMPUTING. R : A Language and Environment for Statistical Computing. 2016. URL : <https://www.r-project.org/>.
165. ALBERT J. LearnBayes : Functions for Learning Bayesian Inference. 2014. URL : <https://cran.r-project.org/web/packages/LearnBayes/index.html>.
166. MARTIN AD, QUINN KM et HEE PARK J. MCMCpack : Markov Chain Monte Carlo in R. *J Statist Soft* 2011 ;42 :42.

167. CUFFE RL. The inclusion of historical control data may reduce the power of a confirmatory study. *Stat Med* 2011 ;30 :1329–38.
168. DEJARDIN D, DELMAR P, WARNE C et al. Use of a historical control group in a noninferiority trial assessing a new antibacterial treatment : A case study and discussion of practical implementation aspects. *Pharm Stat* 2018 ;17 :169–181.
169. VAN ROSMALEN J, DEJARDIN D, VAN NORDEN Y et al. Including historical data in the analysis of clinical trials : Is it worth the effort? *Stat Methods Med Res* 2018 ;27 :3167–3182.
170. WIJEYSUNDERA D, AUSTIN P, HUX J et al. Bayesian statistical inference enhances the interpretation of contemporary randomized controlled trials. *J Clin Epidemiol* 2009 ;62 :13–21.e5.
171. QUINTANA M, VIELE K et LEWIS RJ. Bayesian Analysis : Using Prior Information to Interpret the Results of Clinical Trials. *JAMA* 2017 ;318 :1605–1606.
172. WHELAN J, LE DELEY MC, DIRKSEN U et al. High dose chemotherapy with busulfan-melphalan and blood autologous stem cell rescue compared to standard chemotherapy in localized high-risk Ewing sarcoma : Results of R2Loc randomization in Euro-E.W.I.N.G.99 and Ewing2008 controlled trials. *J Clin Oncol* 2018. (in press).
173. GRILL J, MASSIMINO M, BOUFFET E et al. Phase II, Open-Label, Randomized, Multicenter Trial (HERBY) of Bevacizumab in Pediatric Patients With Newly Diagnosed High-Grade Glioma. *J Clin Oncol* 2018 ;36 :951–958.
174. THE INTERNATIONAL ADJUVANT LUNG CANCER TRIAL COLLABORATIVE GROUP. Cisplatin-Based Adjuvant Chemotherapy in Patients with Completely Resected Non-Small-Cell Lung Cancer. *N Engl J Med* 2004 ;350 :351–360.
175. MOHER D, SCHULZ KF et ALTMAN DG. The CONSORT statement : revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet* 2001 ;357 :1191–4.

176. NEELON B et O'MALLEY A. The Use of Power Prior Distributions for Incorporating Historical Data into a Bayesian Analysis. Department of Health Care Policy, Harvard Medical School. 2009.
177. RIETBERGEN C, KLUGKIST I, JANSSEN KJM et al. Incorporation of historical data in the analysis of randomized therapeutic trials. *Contemp Clin Trials* 2011 ;32 :845–55.

Annexe A

Catégories de Web of Science utilisées

Catégories de la base de données Web of science utilisées pour sélectionner les domaines recherche médical et statistiques.

- Anesthesiology
- Cardiac cardiovascular systems
- Clinical neurology
- Critical care medicine
- Emergency medicine
- Gastro enterology Hepatology
- Geriatrics Gerontology
- Hematology
- Infectious diseases
- Medicine general internal
- Medicine research experimental
- Obstetrics gynecology
- Oncology
- Ophthalmology
- Pathology
- Pediatrics
- Peripheral vascular disease
- Pharmacology pharmacy
- Respiratory system
- Rheumatology
- Statistics probability
- Surgery
- Transplantation
- Urology Nephrology

— Annexe B —

Spécialités médicales et journaux considérés
pour la seconde partie de la recherche

*Annexe B. Spécialités médicales et journaux considérés
pour la seconde partie de la recherche*

Spécialités médicales	Journaux	Facteurs d'impact	Nombre d'articles	
		2014	Identifiés ¹	Inclus ²
Journaux généraux	New England Journal of Medicine	55.87	2	0
	The Lancet	39.21	Irréalisable ³	-
	JAMA	35.30	Irréalisable	-
	British Medical Journal	17.45	5	0
	PlosOne Medicine	14.43	2	0
	BMC-Medicine	7.36	2	0
	Canadian Medical Association Journal	5.96	1	0
Cardiologie	Journal of the American college of cardiology	16.50	9	0
Dermatologie	Journal of investigative dermatology	7.22	0	0
Médecine d'urgence	Annals of emergency medicine	4.70	Irréalisable	-
	Resuscitation	4.17	Irréalisable	-
	Shock	3.05	2	0
Endocrinologie	The Lancet diabetes and endocrinology	9.19	Irréalisable	-
	Diabetes Care	8.10	3	0
Gastro-enterologie	Gastro-enterology	16.72	Irréalisable	-
	Gut	14.66	0	0
Gérontologie	The journal of gerontology : series a	5.42	0	0
Hématologie	Blood	10.45	8	0
Hépatologie	Hepatology	11.06	0	0
Maladies infectieuses	The Lancet infectious diseases	19.45	Irréalisable	-
	Clinical infectious diseases	9.34	2	0
Néphrologie	Journal of the American society of nephrology	9.34	1	0
Neurologie	The Lancet neurology	21.82	Irréalisable	-
	Annals of neurology	9.98	0	0
Obstétrique	Human reproduction update	10.17	0	0
	Obstetrics and Gynecology	5.18	0	0

Spécialités médicales	Journaux	Facteurs d'impact	Nombre d'articles	
		2014	Identifiés ¹	Inclus ²
Oncologie ⁴	The Lancet oncology	24.69	Irréalizable	-
	Journal of Clinical Oncology	18.43	62	10
	Journal of the National Cancer Institute	12.58	12	1
	Clinical Cancer Research	8.72	20	0
	Annals of Oncology	7.04	20	0
	European Journal of Cancer	5.42	Irréalizable	-
	Cancer	5.07	22	3
Ophtalmologie	Lung Cancer	3.96	Irréalizable	-
	Progress in retinal and eye research	8.73	Irréalizable	-
	Ophtalmology	6.14	Irréalizable	-
	American Journal of ophtalmology	3.87	Irréalizable	-
Otorhinolaryngologie	Investigate Ophtalmology and visual science	3.40	0	0
	Ear and hearing	2.84	0	0
Rhumatologie	Annals of the Rheumatic Diseases	10.38	0	0
Transplantation	Journal of Heart and lung transplantation	6.65	Irréalizable	-
	American Journal of transplantation	5.68	1	0
Urologie	European Urology	13.94	0	0
Total			174	14

Spécialités médicales et journaux considérés pour la seconde partie de la recherche

¹ Articles identifiés par la recherche réalisée sur le site des 28 journaux sélectionnés exceptés les articles déjà trouvé par la recherche systématique.

² Article retenus après examen du texte complet afin de vérifier leur éligibilité.

³ Recherche dans le texte complet des articles impossible dans ces journaux.

⁴ Nous avons considéré un grand nombre de journaux dans le domaine de l'oncologie car la plupart des articles identifiés par la première partie de la recherche rapporté des essais en oncologie.

— Annexe C —

Formulaire d'extraction de données

Supplementary Materials 3: Data extraction form

1. Reference of the article (1st author year of publication)
2. Identified from: PubMed WoS Journal Website Clinicaltrials.gov
3. Are the study results published in a peer-reviewed journal? Yes No
4. Is the trial protocol publically available? On clinicatrial.gov another trial registry Published in a journal No details
Reference of the trial

5. **Article retained:** Yes (Bayesian survival analysis in a clinical trial) No

6. **If not, type of article => STOP**

- 6.1. No Bayesian analysis
- 6.2. Clinical trial with a Bayesian analysis of an endpoint other than TTE (CRM, Bayesian adaptive randomisation or monitoring based on a binary outcome, PK-PD...) , please specify
- 6.3. No clinical trial (observational study...) , please specify
- 6.4. Meta-analysis , please specify
- 6.5. Methodological only study , please specify

7. **If yes to Q5**, does the manuscript describe a:

Clinical trial	Re-analysis of a completed trial for a clinical purpose (same objective as the original trial)	Re-analysis of a completed trial for a clinical purpose (different objective from the original trial)	Re-analysis of a completed trial to illustrate a new methodology
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

8. If it is a re-analysis, are the primary study results published in a peer-reviewed journal? Yes No Year of pub.....

9. Is the **phase** of the trial specified in the publication? Yes No

10. Phase:

Phase I Phase I/II Phase II Phase II/III Phase III Phase IV No information

11. What medical field does the study concern?

Oncology	Cardiology	Pneumology	orthopaedics	Gynaecology	Obstetrics	Nephrology	Other
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

If other, please state.....

12. Please specify the **disease** to be treated.....

13. Is it a rare condition? Yes No

14. Disease prevalence..... Not Applicable

15. Is information about disease prevalence/rarity derived from the article Personal search Both

16. **Age** range? Preborn Children/Teenager Adults Children/teenager to adults Not Known

17. **Accrual** from year to year Not Known

18. The trial is **evaluating/comparing**

Medical device Drugs Radiation Surgery Observation Other

19. Is the control arm the current standard of care? Yes No No control arm Not specified

20. If drug(s) is (are) evaluated: are treatments combinations of drugs? Yes No Not known

21. What is (are) the name(s) of the evaluated drug(s)? Not applicable

22. Is it a novel drug/medical device/treatment? Yes No Not applicable

23. If no, what is the novelty? new formulation new indication new combination new modality
or comparative effectiveness of currently used treatments/strategies ?

24. Is it a **multicentre** trial? Single centre Multicentre National International Multicentre unsp NK

25. Where are centres located? USA Canada UK Italy FR Other EU Asia Other
if other, specify

26. Who was responsible for designing and conducting the current **statistical analysis**?

27. Institution of the first author

28. Institution of the last author

29. **Sponsor** of the current study? Public sector Private sector Not applicable NK

30. Does the current manuscript mention **Bayesian methods**:

in the title in the abstract in the keywords in the text

31. Are the details about Bayesian methods provided? in the full text in Appendix Both None

32. Are Bayesian results presented in the abstract? Yes No

Original trial design

33. **Blind** Single Blind Double Blind Open-label Not known
34. **Randomisation level** Cluster randomisation Individually randomised Non randomised Not known
35. Randomisation Fixed allocation ratio Adaptive allocation ratio Not Known Not applicable
36. **Trial design** Parallel group Cross-over Factorial design Single arm Other

If other, specify.....

37. If an **adaptive design** is used, what type(s) of adaptation(s) is (are) considered?

Pre-planned group sequential monitoring	Sample size re-estimation based on nuisance parameter	Sample size adaptation based on ttt effect estimate	Interim subgroup selection	Interim ttt selection, MAMS, Pick the winner...	Other	Non adaptive
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

If other, specify.....

38. Number of planned **interim looks** Not specified
39. **Main hypothesis/objective** Efficacy trial Non-inferiority trial/equivalence Other Not specified
40. How was the **sample size** calculated? Frequentist Pragmatic sample size Bayesian method Not known
41. Planned sample size Not known Attained sample size..... Not known
42. Planned number of events Not known Number of events observed..... Not known
43. If the planned sample size > sample size of the published study, was this due to:
 Slow recruitment Early stopping for benefit Early stopping for futility Early stopping for safety Not Known
44. What is the **primary endpoint of the original trial**?
 OS PFS EFS DFS RFS Other TTE endpoint Non TTE endpoint
- If other, please specify
45. Was the primary endpoint an efficacy outcome? Yes No
46. Type of statistics to be used for the planned primary analysis of the primary endpoint?
 Frequentist Bayesian Both
47. If trial re-analysis, do the authors use the same primary endpoint? Yes No NA If no, detail.....
48. Type of statistics to be used for the analysis of the primary outcome considered in the current study?
 Frequentist Bayesian Both
49. **Secondary endpoints**?
 OS PFS EFS DFS RFS Other TTE endpoint Non TTE endpoint
- If other, please specify
50. Type of statistics to be used for the primary analysis of the secondary outcome(s)? Frequentist Bayesian Both

Bayesian analysis

51. **Prior to the original trial, was the Bayesian approach pre-specified for the**

Primary analysis	Secondary analysis	Monitoring	Selection variable	Sensitivity analysis	Randomisation	Other	No Bayesian analyses pre-specified
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

If other, state

52. **Regarding the Bayesian survival analysis performed in this paper, do the authors analyse**

- TTE endpoint** **Dichotomised TTE endpoint** **Another endpoint**

53. Do the authors justify the use of Bayesian statistics? Yes No
- If yes, specify
54. What Bayesian methods are referenced?.....
55. Was a **minimum important efficacy effect** pre-specified (which may, or may not, be used for final decision making)?
 Yes No Not stated Please detail.....
56. If yes, is this MID expressed in terms of Absolute ttt effect Relative ttt effect Both Not stated
57. If yes to Q55, were sample size calculations performed using this MID?
 Yes No NA (pragmatic sample size) No Known
58. Was a final **decision rule pre-specified for the trial** based on the posterior or predictive distribution?
 Yes No Not stated
- Please detail.....
59. Was the final decision rule calibrated to control frequentist error rates at nominal values? Yes No NA
60. Were prior investigations of the trial operating characteristics conducted? Yes No NA

If Bayesian analysis of a dichotomised TTE endpoint is performed (Q52)

61. How is the **prior distribution** defined?

Quantity e.g. treatment effect	Details (including enthusiastic, sceptical, neutral...)	Parameter e.g. delta	Details of the distribution e.g. Normal(1,0.5)	Historical data	Expert opinion	Reference prior	Non-informative prior	Not Known
<input type="checkbox"/>	Baseline hazard			<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	Parameter in the experimental arm			<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	Absolute treatment effect (1)			<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	Absolute ttt effect (if several priors) (2)			<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	Absolute ttt effect (if several priors) (3)			<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	Relative treatment effect (1)			<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	Relative ttt effect (if several priors) (2)			<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	Relative ttt effect (if several priors) (3)			<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	Covariate effect			<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	Other			<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	Unspecified							

62. Are multiple types of prior distributions considered regarding parameter of interest? Yes No

63. If yes to Q62 do the prior distributions cover a range of opinions on the benefit of the experimental treatment? Yes No

Methods: Bayesian analysis of original TTE endpoint (Q52)

64. What **modelling** approach is used to analyse the survival data?

Parametric	Semiparametric	Nonparametric	Other	Not known
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

If other, please state

65. If **parametric modelling** is used, which model is the data assumed to follow?

1-paramater Exponential	Piecewise Exponential	Weibull	Gamma	Log normal	Log logistic	Cure model	Other	Not stated
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

If other, specify

66. Do the authors justify the model used? Yes No

67. If yes to Q66, how do they justify it?.....

68. If it is a comparative trial, do the authors assume **proportional hazards** for the treatment effect? Yes No Not known

69. How is the **prior distribution** defined?

Quantity	Details (including enthusiastic, sceptical, neutral...)	Parameter e.g. ln(HR)	Details of the distribution e.g. Normal(1,0.5)	Historical data	Expert opinion	Reference prior	Non-informative prior	Not Known
<input type="checkbox"/> Control group model parameters				<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> Experimental arm model parameters				<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> Absolute treatment effect (1)				<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> Absolute ttt effect (if several priors) (2)				<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> Absolute ttt effect (if several priors) (3)				<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> Relative treatment effect (1)				<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> Relative ttt effect (if several priors) (2)				<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> Relative ttt effect (if several priors) (3)				<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> Covariates effect				<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> Other				<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> Unspecified				<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

70. Total number of parameters of the model..... Not specified

71. Number of parameters characterising the treatment effect? Not specified

72. Are several prior distributions considered for the treatment effect?.....Yes No

73. If yes to Q72 do the prior distributions cover a range of opinions on the benefit of the experimental treatment? Yes No

Annexe C. Formulaire d'extraction de données

74. **Historical data:** Do priors incorporate historical data? Yes No Not known
75. Do the authors provide references for the historical data? Yes No
- Please detail
76. From where are the historical data derived?
 Randomised controlled trials Other clinical study Case reports Not Known
77. From how many randomised trials/studies are data synthesized? Not stated
78. How many historical data (number of patients) are synthesized? Not stated
79. How do they do this?

Bayesian random effects meta-analysis	Power prior	Commensurate prior	Bayesian mixture model	Pooled datasets	Other	Not known
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

If other, please state

80. Are the historical data shifted? Yes No Not known
81. Are the historical data down-weighted? Yes No Not known
82. If yes to Q80 or Q81, are the historical data shifted/down-weighted
 Non-adaptively (/pre-specified approach) Adaptively according to their commensurability with the new trial data?
83. If yes to Q80 or Q81, which method(s) are used for combine/shift/down-weight the historical data?
 Tan & Machin (BMJ 2003) Ibrahim & Chen (Springer 2001) Other
- If other, please specify

84. **Expert opinions:** Do priors incorporate expert opinion? Yes No Not known
85. How were experts selected? Not stated
86. How many experts were interviewed? Not stated
87. How were opinions of different experts synthesized?
 Delphi Behavioural aggregation Mathematical aggregation Not stated
88. What methods are cited for the expert opinion prior elicitation? Not reported

89. Overall, are prior distributions categorised into
 Sceptical /Neutral /Enthusiastic priors / Not subdivided

90. How is (are) the **prior distribution(s) reported?**
 Definition of the distribution Graphical representations Measure of location Credibility intervals
 Quantiles of the distribution Effective sample size Other Not reported
- If other, please state

Software

91. Which software/language do the authors use to perform the Bayesian analysis?

SAS	R	WinBugs	OpenBugs	Stan	Fortran	Other	Not stated
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

If other, state

Results

92. Type of statistics used for the primary analysis of this endpoint? Bayesian Bayesian + frequentist

93. In the results, is the Bayesian analysis mentioned as part of

Primary analysis	Secondary analysis	Post-hoc analysis	Monitoring	Selection variable	Sensitivity analysis	Randomisation	Other
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

If other, stated

94. How is (are) the **posterior distribution(s) of the treatment effect reported?**

- Definition of the distribution Graphical representations Measure of location Credibility intervals
 Highest Posterior Density intervals Quantiles of the distribution Posterior probabilities of exceeding a value
 Predictive probabilities of exceeding a value Not reported Other

95. Do the authors provide graphical comparisons of the prior, posterior and likelihood distributions?

- Prior and Posterior Prior, Posterior and Likelihood None

96. Do they perform a **sensitivity analysis** of the prior distribution? Yes No

97. If the relative treatment effect is modelled, are the results presented in terms of

- Relative treatment effect Absolute treatment effect Both None Not specified

Please specify

98. What was the **conclusion** of the Bayesian analysis?

99. Was the Bayesian criterion (from **QErreur ! Source du renvoi introuvable.**) for success (i.e., efficacy/progression to the next stage of development) satisfied?

- Yes No Not Applicable

100. If a frequentist analysis was performed, was the frequentist criterion for success met? Yes No Not Applicable

— Annexe D —

Détails de l'analyse Bayésienne utilisée
par les 28 articles inclus
dans la revue de la littérature

Tableau D1. Détails de l'analyse Bayésienne des 28 articles inclus

Article	Design	Utilisation du Bayésien	Modélisation Bayésienne	Prior du paramètre d'intérêt	Règle de décision pour l'analyse intermédiaire	Règle de décision pour l'analyse finale
Anderson 1994	Simple bras	Analyse secondaire	Non-spécifié	Non-spécifié	-	Non-spécifié
Chamberlain 2008	Simple bras	Monitoring (1 analyse intermédiaire) + Analyse finale	Survie à un temps fixé	Non-spécifié	Futilité : P(PFS à 6 mois > 0.6) < 10% Efficacité : P(PFS à 6 mois > 0.6) > 40% (Probabilité prédite)	Futilité : P(PFS à 6 mois > 0.7) < 10% Efficacité : P(PFS à 6 mois > 0.7) > 40%
Curtis 2013	Groupes parallèles	Monitoring (2 analyses intermédiaires) + Analyse finale	Semi-paramétrique (Processus exponentiel par morceaux)	Non-informatif	Futilité : P(HR > 0.9) > 90% Succès : P(Essais simulés concluant au succès) > 99% (Probabilité prédite)	P(HR < 1) > 97.5% (Probabilité a posteriori)
De Lima 2010	Simple bras	Analyse secondaire	Log-normal	Non-informatif	-	P($\beta > 0$) (Probabilité a posteriori)
Estey 2002	Simple bras	Monitoring (nombre d'analyses non-spécifié)	Survie à un temps fixé	Non-spécifié	P(Réduction de 0.15 du taux de mortalité précoce)	-
Garcia-Manero 2012	Simple bras	Monitoring (3 analyses intermédiaires)	Exponentiel à un paramètre	Non-spécifié	Futilité : P(PFS médiane \geq 7 mois) < 2% (Probabilité a posteriori)	-
Genari 2006	Groupes parallèles	Monitoring (Analyse intermédiaire non planifiée)	Approche hybride	Données historiques + Extrêmes + Non-informatif	Futilité P(amélioration PFS médiane \geq 3 mois) (Probabilité a posteriori)	-
Hensley 2013	Simple bras	Monitoring (évaluation continue)	Non-spécifié	Non-spécifié	Non-spécifié	-
Kanakry 2014	Simple bras	Analyse secondaire	Log-normal	Non-informatif	-	P($\beta > 0$) (Probabilité a posteriori)

Article	Design	Utilisation du Bayésien	Modélisation Bayésienne	Prior du paramètre d'intérêt	Règle de décision pour l'analyse intermédiaire	Règle de décision pour l'analyse finale
Le Deley 2014	Groupes parallèles	Analyse secondaire	Approche hybride	Données historiques	-	P(HR > 1.43) P(HR > 1.25) (Probabilité a posteriori)
Levitt 1996	Groupes parallèles	Analyse post-hoc	Survie à un temps fixé	Données historiques	-	P(Différence entre les traitements) (Probabilité a posteriori)
Lu 2010	Groupes parallèles	Monitoring (Analyse intermédiaire non planifiée)	Exponentiel à un paramètre	Non-spécifié	$P(\lambda_1 > \lambda_2) > 97.5\%$ (Probabilité prédite)	-
Maki 2007	Groupes parallèles	Analyse principale	Log-normal	Non-informatif	-	$P(\beta > 0)$ (Probabilité a posteriori)
Ménard 2001	Groupes parallèles	Analyse post-hoc	Semi-paramétrique	Non-spécifié	-	$P(HR > 1)$ (Probabilité a posteriori)
Miksad 2009	Groupes parallèles	Analyse post-hoc	Approche hybride	Données historiques + Opinions d'experts + Non-informatif	-	P(bénéfice de survie) (Probabilité a posteriori)
Morris 2013	Groupes parallèles (Design cross-over)	Analyse secondaire	Semi-paramétrique	Non-informatif	-	$P(HR > 1)$ (Probabilité a posteriori)
Motzer 2014	Groupes parallèles	Analyse principale	Approche hybride	Non-informatif	-	$P(HR \leq 1.1) \geq 50\%$ $P(HR \leq 1.27) \geq 90\%$ (Probabilité a posteriori)
Muss 2009	Groupes parallèles	Monitoring (4 analyses intermédiaires)	Approche hybride	Non-informatif	Futilité : $P(HR < 0.8046) \geq 80\%$ après 600 patients $P(HR < 0.8046) \geq 70\%$ après 900 patients $P(HR < 0.8046) \geq 60\%$ après 1200 patients Non-inferiorité : $P(HR > 0.8046) \geq 99\%$ (Probabilité prédite)	- -

Article	Design	Utilisation du Bayésien	Modélisation Bayésienne	Prior du paramètre d'intérêt	Règle de décision pour l'analyse intermédiaire	Règle de décision pour l'analyse finale
Norman 2009	Groupes parallèles	Analyse secondaire	Log-normal	Non-informatif	-	$P(\beta > 0)$ (Probabilité a posteriori)
Pecorelli 2009	Groupes parallèles	Monitoring (Analyse intermédiaire non planifiée)	Survie à un temps fixé	Extrêmes + Non-informatif	Futilité : P(Amélioration de 15% de la PFS à 2 ans)	-
Satouchi 2014	Groupes parallèles	Monitoring (Analyse intermédiaire non planifiée)	Approche hybride	Non-spécifié	Non-inferiorité : P(HR < 1.31) (Probabilité prédite)	-
Sharma 2012	Groupes parallèles	Analyse secondaire	Paramétrique non-spécifié	Non-informatif	-	$P(\beta > 0)$ (Probabilité a posteriori)
Shulman 2012	Design factoriel	Monitoring (5 analyses intermédiaires)	Exponentiel à un paramètre	Non-informatif	Efficacité : P(HR = 0.77) > 95% (Probabilité prédite)	-
Shulman 2014	Design factoriel	Monitoring (5 analyses intermédiaires)	Exponentiel à un paramètre	Non-informatif	Non-inferiorité : P(HR > 1.3) < 5% (Probabilité prédite)	-
Stephenson 2014	Groupes parallèles	Analyse principale	Approche hybride	Non-spécifié	-	P(Temps avant progression erlotinib > dinaciclib) (Probabilité a posteriori)
Ueno 2013	Groupes parallèles	Analyse principale	Approche hybride	Non-informatif	-	$P(HR < 1.15)$ (Probabilité a posteriori)
Von Minckwitz 2013	Design factoriel	Monitoring (Analyse intermédiaire non planifiée)	Approche hybride	Non-spécifié	Non-spécifié (Probabilité prédite)	-
Wilber 2010	Groupes parallèles	Monitoring (3 analyses intermédiaires) + Analyse finale	Approche hybride	Non-spécifié	Efficacité : P(Ablation de cathéter > Antiarythmique) > 99% Futilité : P(Ablation de cathéter > Antiarythmique) < 1% (Probabilité prédite)	P(Ablation de cathéter > Antiarythmique) > 98% (Probabilité a posteriori)

PFS : Survie sans progression

HR : Hazard Ratio

β : Coefficient pour une covariable donnée

λ : Risque de base

— Annexe E —

Représentations graphiques
des différents scénarios

Figure E1: Représentations graphiques des scénarios S1 à S12: Pas de conflit en termes de distributions de survie- Distribution Weibull pour les données historiques contrôles et les nouvelles données de l'essai. Les courbes noires représentent l'EFS (Kaplan-Meier) des données historiques contrôles hypothétiques, les courbes en couleur représentent l'EFS dans le groupe contrôle (ligne pleine) et expérimental (ligne en pointillés) du nouvel essai

170

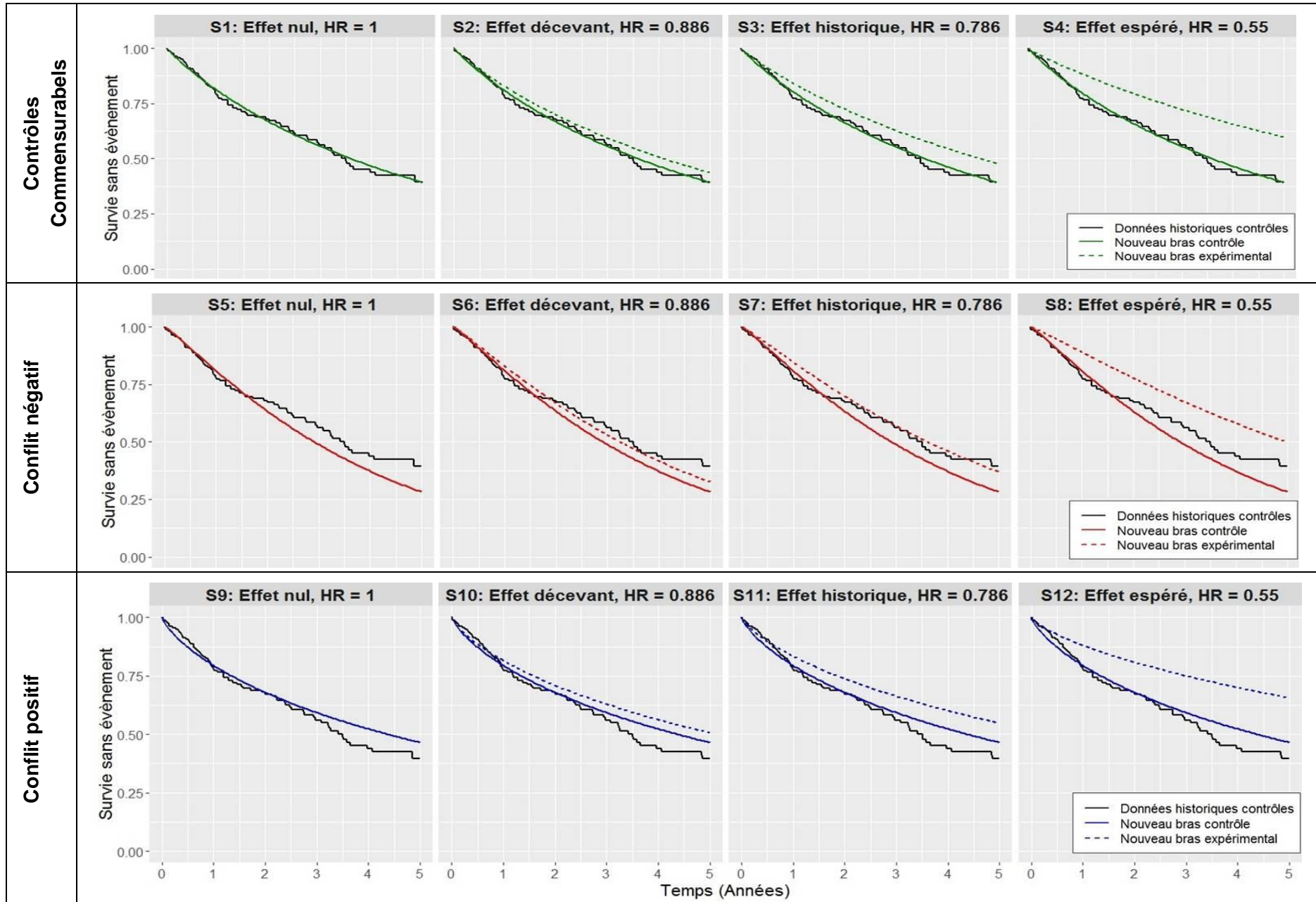


Figure E2: Représentations graphiques des scénarios S13 à S24: Pas de conflit en termes de distributions de survie- Distribution exponentielle par morceaux pour les données historiques contrôles et les nouvelles données de l'essai. Les courbes noires représentent l'EFS (Kaplan-Meier) des données historiques contrôles hypothétiques, les courbes en couleur représentent l'EFS dans le groupe contrôle (ligne pleine) et expérimental (ligne en pointillés) du nouvel essai

171

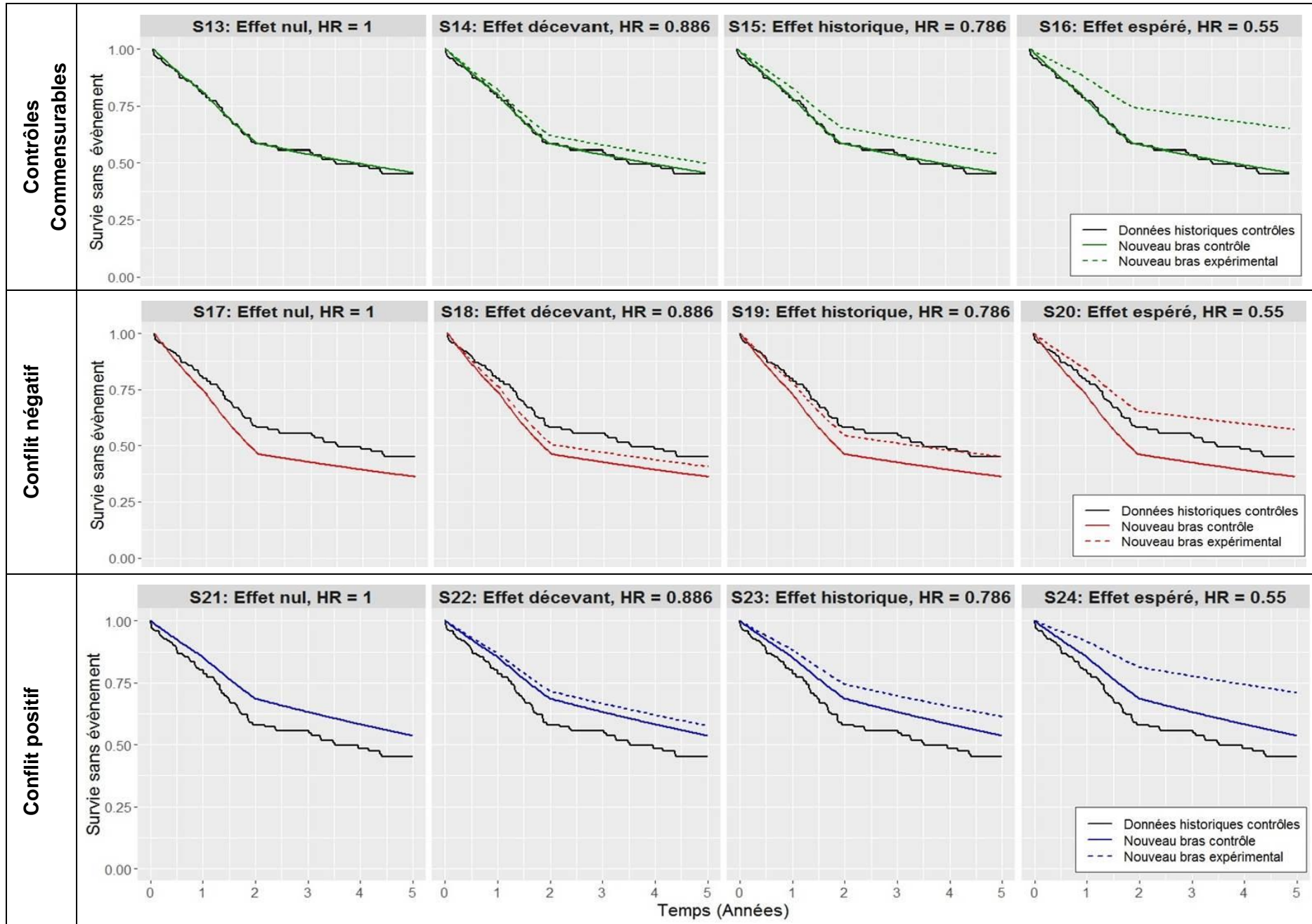
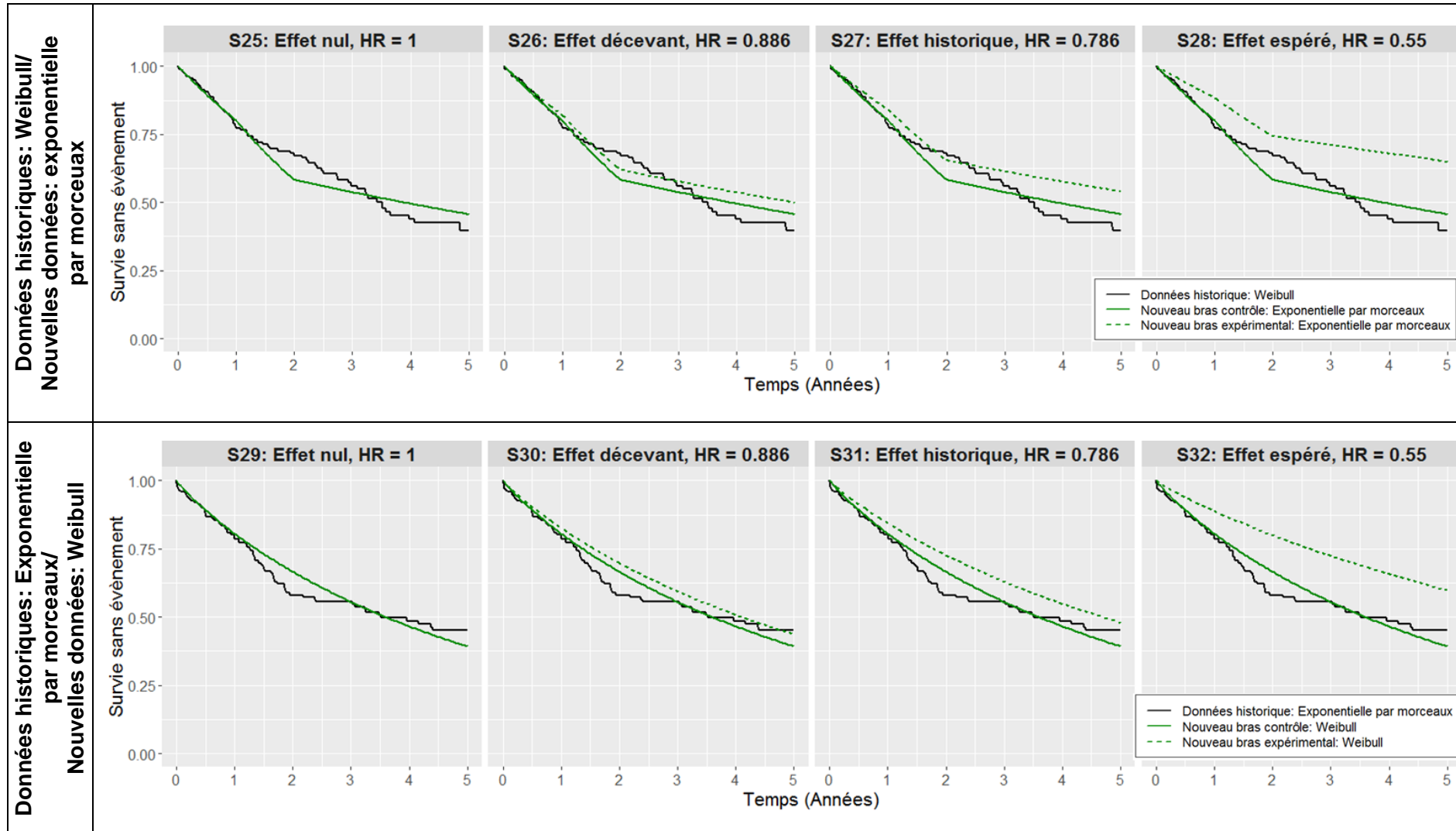


Figure E3: Représentations graphiques des scénarios S25 à S32: Conflit en termes de distribution de survie entre les données historiques et les données du nouvel essai. Les courbes noires représentent l'EFS (Kaplan-Meier) des données historiques contrôles hypothétiques, les courbes en couleur représentent l'EFS dans le groupe contrôle (ligne pleine) et expérimental (ligne en pointillés) du nouvel essai.



— Annexe F —

Impact de conflits en matière de paramètres
entre les données historiques et actuelles

Table F1. Impact de α_0 et ω sur les caractéristiques opératoires pour les scénarios S1 à S4 (Distribution Weibull pour les nouvelles données et les données historiques, bras contrôles commensurables), avec σ_H^2 équivalent à 329 évènements

		Bras contrôles commensurables																
Paramètres	Effet traitement	S1: Effet nul HR = 1				S2: Effet décevant HR = 0.886				S3: Effet historique HR = 0.786				S4: Effet espéré HR = 0.55				
		Biais	SD	RMSE	Alpha	Biais	SD	RMSE	Puissance	Biais	SD	RMSE	Puissance	Biais	SD	RMSE	Puissance	
174	0	0	0.0024	0.263	0.262	0.100	0.0015	0.268	0.267	0.208	0.0022	0.275	0.275	0.359	-0.0075	0.304	0.298	0.794
		0.1	-0.0490	0.240	0.234	0.216	-0.0275	0.234	0.213	0.368	0.0014	0.236	0.210	0.545	0.0582	0.288	0.289	0.900
		0.2	-0.0794	0.222	0.223	0.323	-0.0439	0.210	0.185	0.491	0.0008	0.210	0.176	0.655	0.0999	0.272	0.289	0.944
		0.4	-0.1189	0.195	0.213	0.499	-0.0645	0.178	0.155	0.668	0.0007	0.176	0.132	0.808	0.1573	0.244	0.292	0.983
		0.6	-0.1469	0.171	0.210	0.641	-0.0782	0.155	0.136	0.796	0.0007	0.152	0.102	0.893	0.2015	0.217	0.298	0.993
		0.8	-0.1718	0.147	0.208	0.780	-0.0894	0.133	0.121	0.897	0.0007	0.131	0.074	0.958	0.2432	0.185	0.304	0.997
		1	-0.2035	0.105	0.207	0.969	-0.1017	0.106	0.109	0.989	0.0006	0.107	0.039	0.997	0.3109	0.110	0.313	1
		0.3	0	-0.0058	0.232	0.214	0.084	-0.0071	0.238	0.220	0.211	-0.0070	0.245	0.231	0.396	-0.0158	0.274	0.259
	0.1		-0.0598	0.213	0.196	0.204	-0.0370	0.205	0.171	0.395	-0.0044	0.206	0.170	0.613	0.0542	0.263	0.260	0.957
	0.2		-0.0900	0.197	0.191	0.319	-0.0526	0.183	0.150	0.528	-0.0029	0.183	0.139	0.721	0.0971	0.251	0.264	0.981
	0.4		-0.1268	0.174	0.190	0.499	-0.0705	0.156	0.128	0.706	-0.0014	0.154	0.103	0.859	0.1552	0.227	0.274	0.995
	0.6		-0.1519	0.153	0.192	0.643	-0.0815	0.136	0.116	0.828	-0.0009	0.134	0.078	0.929	0.1990	0.203	0.283	0.999
	0.8		-0.1726	0.133	0.195	0.784	-0.0899	0.121	0.109	0.917	-0.0007	0.119	0.058	0.969	0.2402	0.174	0.292	1
	1		-0.1959	0.102	0.200	0.954	-0.0977	0.103	0.105	0.987	-0.0006	0.103	0.039	0.997	0.3023	0.105	0.305	1
	0.6		0	-0.0095	0.219	0.200	0.082	-0.0108	0.225	0.206	0.222	-0.0111	0.233	0.218	0.418	-0.0199	0.261	0.248
		0.1	-0.0629	0.202	0.186	0.210	-0.0401	0.193	0.161	0.408	-0.0066	0.195	0.159	0.625	0.0501	0.253	0.252	0.973
		0.2	-0.0920	0.187	0.184	0.321	-0.0549	0.173	0.141	0.547	-0.0046	0.173	0.130	0.752	0.0925	0.242	0.258	0.988
		0.4	-0.1275	0.165	0.185	0.500	-0.0715	0.147	0.122	0.725	-0.0023	0.146	0.096	0.877	0.1500	0.221	0.268	0.997
		0.6	-0.1512	0.147	0.188	0.644	-0.0815	0.130	0.113	0.840	-0.0012	0.129	0.074	0.938	0.1935	0.198	0.278	0.999
		0.8	-0.1707	0.129	0.191	0.785	-0.0891	0.116	0.107	0.921	-0.0008	0.115	0.056	0.971	0.2344	0.171	0.287	1
		1	-0.1920	0.101	0.196	0.944	-0.0959	0.101	0.104	0.985	-0.0005	0.102	0.040	0.997	0.2973	0.103	0.300	1
		1	0	-0.0121	0.210	0.193	0.087	-0.0137	0.216	0.200	0.233	-0.0141	0.223	0.212	0.448	-0.0230	0.252	0.242
	0.1		-0.0640	0.194	0.182	0.215	-0.0418	0.185	0.156	0.425	-0.0086	0.187	0.154	0.653	0.0459	0.246	0.249	0.980
	0.2		-0.0923	0.180	0.180	0.325	-0.0558	0.166	0.138	0.559	-0.0059	0.166	0.126	0.771	0.0878	0.237	0.255	0.992
	0.4		-0.1271	0.160	0.182	0.501	-0.0719	0.142	0.121	0.733	-0.0030	0.141	0.094	0.884	0.1443	0.217	0.266	0.998
	0.6		-0.1497	0.143	0.185	0.641	-0.0812	0.126	0.111	0.841	-0.0012	0.125	0.072	0.939	0.1874	0.196	0.274	0.999
	0.8		-0.1685	0.126	0.189	0.776	-0.0883	0.114	0.106	0.920	-0.0009	0.113	0.056	0.973	0.2285	0.169	0.283	1
	1		-0.1892	0.099	0.194	0.933	-0.0945	0.100	0.103	0.983	-0.0006	0.100	0.042	0.997	0.2931	0.101	0.296	1

Table F2. Impact de α_0 et ω sur les caractéristiques opératoires pour les scénarios S5 à S8 (Distribution Weibull pour les nouvelles données et les données historiques, conflit négatif), avec σ_H^2 équivalent à 329 évènements

		Bras contrôles: Conflit négatif																
Paramètres	Effet traitement	S5: Effet nul HR = 1				S6: Effet décevant HR = 0.886				S7: Effet historique HR = 0.786				S8: Effet espéré HR = 0.55				
		Biais	SD	RMSE	Alpha	Biais	SD	RMSE	Puissance	Biais	SD	RMSE	Puissance	Biais	SD	RMSE	Puissance	
175	α_0 0	ω 0	0.0003	0.241	0.243	0.104	0.0008	0.240	0.243	0.224	0.0036	0.245	0.247	0.382	-0.0024	0.263	0.264	0.849
		0.1	-0.0494	0.221	0.222	0.215	-0.0290	0.208	0.195	0.399	0.0030	0.208	0.187	0.580	0.0594	0.254	0.263	0.936
		0.2	-0.0791	0.206	0.214	0.321	-0.0454	0.188	0.172	0.523	0.0025	0.185	0.155	0.693	0.0991	0.244	0.266	0.968
		0.4	-0.1168	0.181	0.208	0.495	-0.0651	0.160	0.147	0.692	0.0019	0.156	0.118	0.832	0.1541	0.222	0.274	0.989
		0.6	-0.1429	0.161	0.206	0.646	-0.0781	0.140	0.132	0.810	0.0014	0.136	0.093	0.904	0.1966	0.200	0.282	0.997
		0.8	-0.1658	0.141	0.205	0.774	-0.0882	0.122	0.120	0.899	0.0013	0.119	0.071	0.957	0.2374	0.172	0.291	0.999
		1	-0.1980	0.103	0.203	0.947	-0.0994	0.100	0.108	0.986	0.0013	0.100	0.042	0.995	0.3024	0.102	0.305	1
		0.3	0	0.0798	0.216	0.213	0.038	0.0755	0.217	0.213	0.115	0.0619	0.223	0.221	0.265	0.0813	0.243	0.239
	0.1		0.0184	0.211	0.195	0.104	0.0289	0.195	0.171	0.261	0.0598	0.188	0.163	0.470	0.1419	0.226	0.250	0.937
	0.2		-0.0180	0.202	0.189	0.181	0.0029	0.179	0.150	0.389	0.0485	0.167	0.135	0.608	0.1772	0.212	0.260	0.971
	0.4		-0.0650	0.185	0.183	0.328	-0.0280	0.156	0.126	0.581	0.0359	0.142	0.102	0.772	0.2214	0.187	0.276	0.992
	0.6		-0.0996	0.166	0.181	0.481	-0.0475	0.138	0.111	0.718	0.0286	0.125	0.080	0.865	0.2522	0.165	0.288	0.998
	0.8		-0.1306	0.145	0.180	0.640	-0.0629	0.121	0.100	0.833	0.0234	0.112	0.063	0.935	0.2781	0.142	0.299	1
	1		-0.1739	0.100	0.179	0.878	-0.0788	0.099	0.089	0.965	0.0189	0.099	0.045	0.999	0.3099	0.100	0.312	1
	0.6		0	0.1150	0.205	0.216	0.020	0.1124	0.206	0.216	0.075	0.1169	0.212	0.224	0.206	0.1212	0.233	0.245
		0.1	0.0548	0.206	0.198	0.067	0.0598	0.191	0.176	0.193	0.0871	0.181	0.169	0.406	0.1781	0.212	0.258	0.931
		0.2	0.0178	0.201	0.189	0.126	0.0304	0.178	0.154	0.313	0.0710	0.163	0.142	0.548	0.2094	0.196	0.270	0.964
		0.4	-0.0327	0.188	0.180	0.251	-0.0051	0.157	0.128	0.496	0.0532	0.140	0.108	0.728	0.2471	0.171	0.285	0.990
		0.6	-0.0709	0.172	0.174	0.385	-0.0283	0.140	0.110	0.648	0.0426	0.124	0.087	0.833	0.2720	0.151	0.297	0.998
		0.8	-0.1067	0.151	0.171	0.547	-0.0472	0.123	0.095	0.775	0.0352	0.111	0.069	0.911	0.2921	0.131	0.306	0.999
		1	-0.1607	0.099	0.166	0.805	-0.0672	0.098	0.079	0.934	0.0287	0.098	0.050	0.984	0.3147	0.099	0.317	1
		1	0	0.1407	0.196	0.224	0.014	0.1395	0.198	0.225	0.056	0.1448	0.204	0.233	0.168	0.1503	0.225	0.255
	0.1		0.0838	0.202	0.205	0.046	0.0848	0.188	0.186	0.145	0.1083	0.177	0.180	0.355	0.2036	0.202	0.269	0.922
	0.2		0.0474	0.200	0.196	0.093	0.0535	0.178	0.164	0.257	0.0888	0.161	0.153	0.499	0.2314	0.185	0.280	0.961
	0.4		-0.0041	0.190	0.183	0.193	0.0150	0.159	0.135	0.429	0.0676	0.139	0.119	0.677	0.2639	0.161	0.294	0.989
	0.6		-0.0448	0.176	0.175	0.311	-0.0113	0.143	0.115	0.582	0.0545	0.124	0.096	0.792	0.2846	0.142	0.304	0.997
	0.8		-0.0851	0.156	0.166	0.458	-0.0330	0.126	0.096	0.723	0.0450	0.112	0.077	0.883	0.3011	0.124	0.312	0.999
	1		-0.1501	0.098	0.156	0.728	-0.0577	0.097	0.071	0.900	0.0368	0.097	0.055	0.971	0.3188	0.098	0.321	1

Table F3. Impact de α_0 et ω sur les caractéristiques opératoire pour les scénarios S9 à S12 (Distribution Weibull les nouvelles données et les données historiques, conflit positif), avec σ_H^2 équivalent à 329 évènements

Bras contrôles: Conflit positif																		
Effet traitement		S9: Effet nul HR = 1				S10: Effet décevant HR = 0.886				S11: Effet historique HR = 0.786				S12: Effet espéré HR = 0.55				
Paramètres		Biais	SD	RMSE	Alpha	Biais	SD	RMSE	Puissance	Biais	SD	RMSE	Puissance	Biais	SD	RMSE	Puissance	
α_0	ω																	
	0	0	0.0009	0.281	0.280	0.102	-0.0003	0.280	0.284	0.201	0.0008	0.286	0.291	0.354	-0.0089	0.315	0.315	0.764
	0.1	-0.0498	0.256	0.248	0.213	-0.0301	0.244	0.227	0.354	0.0007	0.246	0.223	0.538	0.0447	0.300	0.300	0.882	
	0.2	-0.0801	0.236	0.233	0.326	-0.0463	0.219	0.198	0.486	0.0005	0.219	0.185	0.662	0.0872	0.285	0.295	0.931	
	0.4	-0.1205	0.205	0.221	0.494	-0.0669	0.183	0.164	0.684	0.0004	0.181	0.139	0.816	0.1475	0.256	0.294	0.976	
	0.6	-0.1493	0.179	0.215	0.648	-0.0808	0.157	0.143	0.801	0.0003	0.153	0.106	0.902	0.1947	0.227	0.297	0.994	
	0.8	-0.1744	0.153	0.212	0.791	-0.0923	0.133	0.127	0.897	0	0.130	0.076	0.959	0.2402	0.192	0.301	0.999	
	1	-0.2075	0.107	0.211	0.975	-0.1048	0.102	0.112	0.992	0	0.102	0.038	0.998	0.3125	0.104	0.315	1	
0.3	0	-0.0605	0.246	0.237	0.132	-0.0640	0.248	0.244	0.274	-0.0616	0.256	0.250	0.481	-0.0817	0.287	0.286	0.905	
	0.1	-0.1043	0.218	0.214	0.278	-0.0799	0.210	0.192	0.472	-0.0557	0.217	0.191	0.686	-0.0180	0.285	0.279	0.969	
	0.2	-0.1285	0.198	0.207	0.402	-0.0877	0.186	0.168	0.620	-0.0447	0.192	0.157	0.801	0.0287	0.277	0.277	0.987	
	0.4	-0.1577	0.171	0.205	0.591	-0.0968	0.155	0.144	0.783	-0.0315	0.161	0.117	0.910	0.0957	0.258	0.277	0.999	
	0.6	-0.1770	0.149	0.205	0.734	-0.1021	0.135	0.130	0.880	-0.0232	0.138	0.089	0.959	0.1496	0.234	0.280	1	
	0.8	-0.1921	0.130	0.207	0.851	-0.1061	0.118	0.122	0.949	-0.0170	0.120	0.065	0.985	0.2033	0.202	0.284	1	
	1	-0.2083	0.103	0.212	0.975	-0.1094	0.101	0.116	0.995	-0.0110	0.101	0.039	0.999	0.2954	0.102	0.297	1	
	0.6	0	-0.0892	0.231	0.234	0.163	-0.0938	0.235	0.242	0.336	-0.0926	0.243	0.248	0.562	-0.1145	0.275	0.287	0.950
0.1		-0.1272	0.202	0.214	0.331	-0.1012	0.197	0.192	0.550	-0.0769	0.208	0.193	0.763	-0.0529	0.279	0.283	0.986	
0.2		-0.1474	0.183	0.208	0.462	-0.1046	0.175	0.170	0.685	-0.0614	0.185	0.160	0.853	-0.0068	0.275	0.280	0.996	
0.4		-0.1710	0.158	0.207	0.641	-0.1082	0.147	0.146	0.825	-0.0437	0.156	0.121	0.942	0.0606	0.260	0.278	1	
0.6		-0.1862	0.139	0.208	0.772	-0.1104	0.129	0.134	0.909	-0.0326	0.135	0.093	0.973	0.1169	0.240	0.278	1	
0.8		-0.1980	0.122	0.210	0.879	-0.1120	0.114	0.126	0.961	-0.0241	0.119	0.069	0.991	0.1753	0.209	0.279	1	
1		-0.2098	0.102	0.214	0.976	-0.1131	0.100	0.120	0.995	-0.0163	0.100	0.042	1	0.2850	0.101	0.287	1	
1		0	-0.1108	0.221	0.238	0.194	-0.1162	0.226	0.246	0.400	-0.1158	0.235	0.253	0.628	-0.1391	0.268	0.293	0.971
	0.1	-0.1433	0.192	0.218	0.378	-0.1171	0.190	0.197	0.613	-0.0933	0.202	0.199	0.810	-0.0809	0.274	0.289	0.993	
	0.2	-0.1603	0.174	0.213	0.503	-0.1168	0.168	0.175	0.730	-0.0749	0.181	0.167	0.889	-0.0363	0.272	0.286	0.998	
	0.4	-0.1804	0.150	0.211	0.677	-0.1168	0.142	0.152	0.860	-0.0540	0.154	0.128	0.955	0.0302	0.262	0.282	1	
	0.6	-0.1926	0.132	0.212	0.800	-0.1168	0.125	0.139	0.927	-0.0409	0.135	0.100	0.980	0.0872	0.245	0.280	1	
	0.8	-0.2022	0.118	0.213	0.893	-0.1168	0.112	0.130	0.966	-0.0307	0.118	0.075	0.993	0.1487	0.217	0.277	1	
	1	-0.2116	0.100	0.216	0.973	-0.1165	0.099	0.124	0.996	-0.0211	0.099	0.046	1	0.2761	0.100	0.279	1	

Table F4. Impact de α_0 et ω sur les caractéristiques opératoires pour les scénarios S13 à S16 (Distribution exponentielle par morceaux pour les nouvelles données et les données historiques, bras contrôles commensurables), avec σ_H^2 équivalent à 329 évènements

		Bras contrôles commensurables																
Paramètres	Effet traitement	S13: Effet nul HR = 1				S14: Effet décevant HR = 0.886				S15: Effet historique HR = 0.786				S16: Effet espéré HR = 0.55				
	α_0	ω	Biais	SD	RMSE	Alpha	Biais	SD	RMSE	Puissance	Biais	SD	RMSE	Puissance	Biais	SD	RMSE	Puissance
177	0	0	-0.0018	0.274	0.280	0.104	-0.0050	0.277	0.283	0.211	-0.0029	0.287	0.296	0.333	-0.0139	0.311	0.320	0.770
		0.1	-0.0514	0.249	0.248	0.217	-0.0333	0.242	0.225	0.364	-0.0019	0.248	0.229	0.500	0.0419	0.296	0.304	0.880
		0.2	-0.0813	0.230	0.235	0.328	-0.0491	0.217	0.195	0.489	-0.0018	0.221	0.192	0.639	0.0843	0.281	0.300	0.935
		0.4	-0.1205	0.200	0.222	0.508	-0.0685	0.182	0.162	0.680	-0.0010	0.184	0.146	0.799	0.1441	0.254	0.300	0.979
		0.6	-0.1483	0.175	0.216	0.654	-0.0813	0.156	0.142	0.808	-0.0009	0.157	0.112	0.883	0.1910	0.226	0.303	0.993
		0.8	-0.1733	0.148	0.213	0.794	-0.0919	0.133	0.126	0.902	-0.0007	0.132	0.080	0.953	0.2367	0.192	0.306	0.998
		1	-0.2070	0.102	0.211	0.972	-0.1039	0.102	0.111	0.992	-0.0001	0.103	0.038	0.998	0.3132	0.104	0.315	1
		0.3	0	-0.0227	0.241	0.227	0.095	-0.0251	0.246	0.233	0.225	-0.0251	0.256	0.246	0.385	-0.0439	0.283	0.276
	0.1		-0.0732	0.218	0.206	0.226	-0.0509	0.210	0.181	0.408	-0.0177	0.216	0.183	0.598	0.0280	0.276	0.271	0.951
	0.2		-0.1016	0.200	0.200	0.341	-0.0641	0.187	0.158	0.552	-0.0137	0.191	0.151	0.731	0.0732	0.265	0.273	0.981
	0.4		-0.1368	0.174	0.197	0.527	-0.0791	0.157	0.135	0.734	-0.0089	0.160	0.112	0.862	0.1353	0.242	0.278	0.995
	0.6		-0.1598	0.152	0.198	0.673	-0.0882	0.136	0.122	0.851	-0.0060	0.138	0.085	0.932	0.1833	0.217	0.284	0.998
	0.8		-0.1796	0.132	0.201	0.812	-0.0951	0.119	0.114	0.929	-0.0036	0.120	0.062	0.971	0.2295	0.186	0.291	0.999
	1		-0.2013	0.101	0.205	0.966	-0.1017	0.101	0.109	0.991	-0.0011	0.101	0.038	0.999	0.3027	0.102	0.305	1
	0.6	0	-0.0299	0.227	0.214	0.096	-0.0325	0.233	0.220	0.239	-0.0330	0.243	0.234	0.421	-0.0510	0.272	0.266	0.906
		0.1	-0.0791	0.206	0.196	0.234	-0.0563	0.198	0.172	0.438	-0.0228	0.204	0.173	0.642	0.0209	0.267	0.264	0.970
		0.2	-0.1065	0.189	0.193	0.352	-0.0683	0.176	0.151	0.579	-0.0175	0.181	0.142	0.763	0.0656	0.258	0.267	0.990
		0.4	-0.1394	0.165	0.192	0.534	-0.0815	0.149	0.130	0.752	-0.0116	0.152	0.106	0.881	0.1276	0.237	0.273	0.997
		0.6	-0.1614	0.145	0.195	0.680	-0.0894	0.130	0.120	0.863	-0.0077	0.132	0.081	0.943	0.1757	0.214	0.279	0.999
		0.8	-0.1788	0.127	0.198	0.807	-0.0952	0.116	0.113	0.933	-0.0048	0.117	0.060	0.976	0.2221	0.184	0.286	0.999
		1	-0.1984	0.100	0.203	0.958	-0.1007	0.100	0.108	0.989	-0.0020	0.100	0.039	0.998	0.2973	0.101	0.300	1
	1	0	-0.0345	0.218	0.208	0.104	-0.0376	0.224	0.215	0.255	-0.0381	0.234	0.228	0.454	-0.0555	0.264	0.261	0.925
		0.1	-0.0825	0.197	0.193	0.244	-0.0596	0.190	0.168	0.459	-0.0262	0.197	0.170	0.669	0.0153	0.261	0.261	0.981
		0.2	-0.1088	0.182	0.190	0.357	-0.0706	0.170	0.148	0.597	-0.0202	0.175	0.140	0.783	0.0595	0.253	0.264	0.992
		0.4	-0.1403	0.159	0.191	0.541	-0.0827	0.144	0.129	0.763	-0.0133	0.147	0.105	0.893	0.1207	0.234	0.271	0.997
		0.6	-0.1610	0.141	0.193	0.686	-0.0898	0.127	0.119	0.867	-0.0090	0.129	0.081	0.945	0.1685	0.212	0.277	0.999
		0.8	-0.1779	0.124	0.196	0.807	-0.0951	0.113	0.113	0.934	-0.0056	0.115	0.060	0.977	0.2154	0.183	0.283	0.999
		1	-0.1963	0.099	0.201	0.949	-0.1000	0.099	0.108	0.986	-0.0026	0.099	0.041	0.998	0.2930	0.100	0.295	1

Table F5. Impact de α_0 et ω sur les caractéristiques opératoires pour les scénarios S17 à S20 (Distribution exponentielle par morceaux pour les nouvelles données et les données historiques, conflit négatif), avec σ_H^2 équivalent à 329 évènements

Bras contrôles: Conflit négatif																		
Paramètres	Effet traitement		S17: Effet nul HR = 1				S18: Effet décevant HR = 0.886				S19: Effet historique HR = 0.786				S20: Effet espéré HR = 0.55			
	α_0	ω	Biais	SD	RMSE	Alpha	Biais	SD	RMSE	Puissance	Biais	SD	RMSE	Puissance	Biais	SD	RMSE	Puissance
178	0	0	-0.0068	0.251	0.254	0.109	-0.0046	0.256	0.261	0.219	-0.0111	0.262	0.265	0.368	-0.0079	0.283	0.286	0.819
		0.1	-0.0581	0.230	0.228	0.221	-0.0318	0.223	0.209	0.387	-0.0086	0.224	0.202	0.559	0.0582	0.271	0.280	0.913
		0.2	-0.0878	0.212	0.219	0.336	-0.0470	0.200	0.183	0.515	-0.0074	0.199	0.168	0.688	0.0997	0.257	0.281	0.954
		0.4	-0.1255	0.185	0.211	0.511	-0.0662	0.170	0.154	0.687	-0.0054	0.166	0.127	0.836	0.1558	0.232	0.286	0.984
		0.6	-0.1516	0.162	0.209	0.650	-0.0786	0.148	0.136	0.808	-0.0038	0.144	0.100	0.916	0.1994	0.207	0.292	0.994
		0.8	-0.1740	0.139	0.209	0.793	-0.0894	0.127	0.122	0.896	-0.0025	0.124	0.074	0.963	0.2408	0.177	0.299	0.998
		1	-0.2030	0.101	0.207	0.966	-0.1015	0.101	0.109	0.989	-0.0012	0.101	0.040	0.997	0.3080	0.102	0.310	1
		0.3	0	0.1096	0.224	0.235	0.028	0.1107	0.229	0.242	0.088	0.1041	0.236	0.245	0.216	0.1019	0.258	0.268
	0.1		0.0475	0.221	0.213	0.083	0.0581	0.209	0.196	0.211	0.0760	0.202	0.185	0.397	0.1616	0.236	0.273	0.899
	0.2		0.0092	0.214	0.203	0.148	0.0286	0.194	0.172	0.327	0.0619	0.180	0.154	0.545	0.1949	0.218	0.280	0.947
	0.4		-0.0424	0.197	0.192	0.286	-0.0072	0.170	0.144	0.506	0.0456	0.153	0.118	0.724	0.2364	0.190	0.292	0.980
	0.6		-0.0818	0.179	0.186	0.429	-0.0312	0.150	0.124	0.655	0.0357	0.134	0.093	0.833	0.2651	0.166	0.301	0.992
	0.8		-0.1188	0.155	0.182	0.588	-0.0518	0.131	0.107	0.790	0.0284	0.118	0.072	0.921	0.2895	0.142	0.310	0.998
	1		-0.1724	0.100	0.177	0.865	-0.0751	0.100	0.085	0.954	0.0212	0.100	0.045	0.989	0.3188	0.101	0.321	1
	0.6	0	0.1542	0.216	0.249	0.014	0.1543	0.217	0.255	0.055	0.1473	0.224	0.257	0.164	0.1430	0.248	0.278	0.735
		0.1	0.0954	0.216	0.228	0.051	0.0974	0.206	0.212	0.150	0.1095	0.196	0.200	0.332	0.1962	0.222	0.283	0.884
		0.2	0.0570	0.213	0.216	0.097	0.0644	0.194	0.188	0.251	0.0898	0.177	0.169	0.467	0.2252	0.203	0.290	0.939
		0.4	0.0029	0.202	0.201	0.204	0.0237	0.174	0.157	0.411	0.0671	0.152	0.132	0.653	0.2600	0.175	0.301	0.975
		0.6	-0.0407	0.187	0.189	0.321	-0.0056	0.155	0.133	0.570	0.0532	0.134	0.106	0.782	0.2831	0.154	0.310	0.990
		0.8	-0.0841	0.165	0.179	0.480	-0.0307	0.135	0.110	0.714	0.0425	0.118	0.083	0.884	0.3021	0.132	0.317	0.997
		1	-0.1578	0.099	0.163	0.764	-0.0625	0.099	0.075	0.908	0.0318	0.099	0.052	0.977	0.3237	0.100	0.326	0.999
	1	0	0.1834	0.211	0.264	0.011	0.1828	0.209	0.269	0.041	0.1754	0.216	0.271	0.138	0.1695	0.240	0.289	0.719
		0.1	0.1287	0.211	0.244	0.035	0.1252	0.202	0.228	0.121	0.1328	0.192	0.215	0.291	0.2177	0.212	0.293	0.867
		0.2	0.0921	0.211	0.231	0.072	0.0914	0.193	0.204	0.202	0.1101	0.176	0.185	0.408	0.2435	0.193	0.299	0.928
		0.4	0.0382	0.204	0.214	0.156	0.0475	0.176	0.172	0.356	0.0838	0.153	0.147	0.601	0.2736	0.167	0.309	0.971
		0.6	-0.0072	0.192	0.199	0.256	0.0157	0.160	0.146	0.497	0.0669	0.135	0.120	0.735	0.2934	0.146	0.316	0.987
		0.8	-0.0552	0.172	0.184	0.398	-0.0134	0.140	0.119	0.647	0.0535	0.120	0.094	0.846	0.3093	0.127	0.322	0.996
		1	-0.1470	0.099	0.153	0.683	-0.0533	0.099	0.069	0.865	0.0395	0.099	0.058	0.959	0.3272	0.099	0.330	0.999

Table F6. Impact de α_0 et ω sur les caractéristiques opératoire pour les scénarios S21 à S24 (Distribution exponentielle par morceaux les nouvelles données et les données historiques, conflit positif), avec σ_H^2 équivalent à 329 évènements

Bras contrôles: Conflit positif																			
Effet traitement		S21: Effet nul HR = 1				S22: Effet décevant HR = 0.886				S23: Effet historique HR = 0.786				S24: Effet espéré HR = 0.55					
Paramètres		Biais	SD	RMSE	Alpha	Biais	SD	RMSE	Puissance	Biais	SD	RMSE	Puissance	Biais	SD	RMSE	Puissance		
α_0	ω	0	0	0.0015	0.299	0.307	0.108	-0.0051	0.307	0.316	0.200	-0.0106	0.315	0.323	0.323	-0.0186	0.345	0.349	0.707
		0.1	-0.0481	0.272	0.270	0.203	-0.0303	0.269	0.255	0.308	-0.0082	0.274	0.251	0.478	0.0491	0.324	0.328	0.833	
		0.2	-0.0782	0.250	0.253	0.310	-0.0453	0.242	0.222	0.469	-0.0067	0.244	0.210	0.607	0.0936	0.305	0.320	0.900	
		0.4	-0.1188	0.216	0.236	0.502	-0.0655	0.202	0.182	0.654	-0.0050	0.202	0.157	0.775	0.1555	0.271	0.314	0.962	
		0.6	-0.1487	0.186	0.227	0.659	-0.0801	0.171	0.156	0.788	-0.0036	0.169	0.120	0.881	0.2037	0.237	0.314	0.985	
		0.8	-0.1748	0.156	0.220	0.799	-0.0919	0.143	0.135	0.892	-0.0021	0.140	0.087	0.953	0.2499	0.197	0.316	0.996	
		1	-0.2114	0.103	0.215	0.983	-0.1063	0.104	0.112	0.995	0	0.104	0.035	0.999	0.3212	0.105	0.323	1	
		0.3	0	-0.1517	0.259	0.287	0.214	-0.1585	0.268	0.299	0.392	-0.1618	0.277	0.309	0.574	-0.1670	0.311	0.343	0.926
0.1	-0.1749		0.222	0.255	0.392	-0.1476	0.227	0.240	0.591	-0.1207	0.242	0.244	0.756	-0.0919	0.313	0.327	0.971		
0.2	-0.1871		0.197	0.243	0.537	-0.1413	0.200	0.211	0.722	-0.0973	0.217	0.206	0.853	-0.0408	0.308	0.319	0.989		
0.4	-0.2017		0.165	0.233	0.728	-0.1343	0.166	0.178	0.867	-0.0694	0.183	0.159	0.945	0.0363	0.291	0.310	0.997		
0.6	-0.2107		0.142	0.229	0.848	-0.1299	0.142	0.157	0.936	-0.0503	0.157	0.123	0.978	0.1006	0.266	0.304	0.999		
0.8	-0.2179		0.123	0.228	0.928	-0.1266	0.123	0.143	0.973	-0.0355	0.133	0.090	0.993	0.1683	0.230	0.301	1		
1	-0.2244		0.101	0.228	0.993	-0.1227	0.102	0.128	0.999	-0.0198	0.102	0.042	1	0.2948	0.102	0.297	1		
0.6	0		-0.2058	0.245	0.308	0.301	-0.2122	0.254	0.320	0.506	-0.2147	0.264	0.329	0.700	-0.2190	0.298	0.362	0.965	
	0.1	-0.2145	0.205	0.273	0.510	-0.1860	0.215	0.261	0.713	-0.1611	0.235	0.265	0.854	-0.1505	0.308	0.347	0.990		
	0.2	-0.2191	0.182	0.260	0.648	-0.1724	0.191	0.231	0.822	-0.1312	0.214	0.230	0.922	-0.1012	0.307	0.337	0.997		
	0.4	-0.2241	0.152	0.247	0.810	-0.1560	0.160	0.195	0.926	-0.0951	0.183	0.181	0.974	-0.0249	0.298	0.324	0.999		
	0.6	-0.2274	0.133	0.241	0.901	-0.1464	0.139	0.173	0.965	-0.0714	0.160	0.145	0.990	0.0418	0.280	0.314	1		
	0.8	-0.2295	0.117	0.238	0.953	-0.1388	0.121	0.155	0.987	-0.0512	0.136	0.106	0.998	0.1175	0.247	0.304	1		
	1	-0.2312	0.100	0.234	0.995	-0.1311	0.101	0.136	0.999	-0.0297	0.101	0.047	1	0.2812	0.101	0.283	1		
	1	0	-0.2394	0.235	0.327	0.373	-0.2460	0.245	0.338	0.597	-0.2479	0.255	0.347	0.777	-0.2513	0.290	0.378	0.981	
0.1		-0.2386	0.196	0.289	0.593	-0.2107	0.210	0.279	0.784	-0.1882	0.231	0.285	0.904	-0.1892	0.302	0.364	0.997		
0.2		-0.2381	0.174	0.273	0.722	-0.1921	0.188	0.247	0.876	-0.1551	0.213	0.249	0.951	-0.1439	0.305	0.355	0.998		
0.4		-0.2377	0.146	0.258	0.857	-0.1713	0.159	0.211	0.949	-0.1149	0.186	0.200	0.985	-0.0706	0.300	0.340	1		
0.6		-0.2374	0.128	0.250	0.925	-0.1585	0.139	0.187	0.976	-0.0872	0.162	0.163	0.995	-0.0023	0.286	0.327	1		
0.8		-0.2370	0.114	0.245	0.966	-0.1480	0.121	0.166	0.992	-0.0641	0.139	0.124	1	0.0761	0.259	0.312	1		
1		-0.2363	0.100	0.240	0.996	-0.1375	0.100	0.143	0.999	-0.0372	0.100	0.053	1	0.2706	0.100	0.273	1		

— Annexe G —

Impact des données historiques
pour différentes variances
de la composante informative

Table G1. Impact de α_0 et ω sur les caractéristiques opératoires pour les scénarios S1 à S4 (Distribution Weibull pour les nouvelles données et les données historiques, bras contrôles commensurables), avec σ_H^2 équivalent à 66 évènements.

		Bras contrôles commensurables															
Paramètres	Effet traitement	S1: Effet nul HR = 1				S2: Effet décevant HR = 0.886				S3: Effet historique HR = 0.786				S4: Effet anticipé HR = 0.55			
		Biais	SD	RMSE	Alpha	Biais	SD	RMSE	Puissance	Biais	SD	RMSE	Puissance	Biais	SD	RMSE	Puissance
0	α_0																
	ω																
	0	0.0024	0.263	0.262	0.100	0.0015	0.268	0.267	0.208	0.0022	0.275	0.275	0.359	-0.0075	0.304	0.298	0.794
	0.1	-0.0338	0.246	0.235	0.127	-0.0182	0.239	0.226	0.259	0.0020	0.243	0.225	0.422	0.0381	0.279	0.277	0.835
	0.2	-0.0556	0.234	0.221	0.147	-0.0297	0.225	0.203	0.289	0.0015	0.228	0.199	0.460	0.0706	0.269	0.268	0.854
	0.4	-0.0825	0.218	0.203	0.174	-0.0434	0.207	0.177	0.324	0.0011	0.209	0.168	0.502	0.1142	0.250	0.257	0.879
	0.6	-0.0998	0.205	0.193	0.189	-0.0518	0.195	0.161	0.344	0.0010	0.197	0.149	0.518	0.1455	0.233	0.250	0.889
	0.8	-0.1130	0.193	0.184	0.199	-0.0578	0.186	0.148	0.354	0.0009	0.188	0.135	0.532	0.1719	0.215	0.243	0.895
1	-0.1250	0.181	0.176	0.204	-0.0627	0.178	0.139	0.365	0.0008	0.181	0.124	0.542	0.2014	0.186	0.235	0.900	
0.3	0	-0.0058	0.232	0.214	0.084	-0.0071	0.238	0.220	0.211	-0.0070	0.245	0.231	0.396	-0.0158	0.274	0.259	0.883
	0.1	-0.0405	0.218	0.194	0.114	-0.0255	0.214	0.186	0.265	-0.0068	0.218	0.188	0.469	0.0302	0.258	0.243	0.912
	0.2	-0.0599	0.209	0.184	0.135	-0.0349	0.202	0.169	0.296	-0.0063	0.205	0.167	0.504	0.0625	0.249	0.237	0.930
	0.4	-0.0821	0.195	0.174	0.156	-0.0451	0.188	0.151	0.332	-0.0042	0.191	0.144	0.545	0.1047	0.233	0.229	0.944
	0.6	-0.0954	0.186	0.168	0.169	-0.0509	0.180	0.140	0.353	-0.0030	0.182	0.131	0.565	0.1338	0.217	0.224	0.951
	0.8	-0.1049	0.178	0.164	0.181	-0.0546	0.174	0.134	0.366	-0.0021	0.176	0.122	0.576	0.1580	0.201	0.219	0.954
	1	-0.1124	0.171	0.161	0.187	-0.0573	0.169	0.129	0.376	-0.0014	0.172	0.115	0.588	0.1829	0.177	0.215	0.955
	0.6	0	-0.0095	0.219	0.200	0.082	-0.0108	0.225	0.206	0.222	-0.0111	0.233	0.218	0.418	-0.0199	0.261	0.248
0.1		-0.0424	0.207	0.183	0.114	-0.0277	0.203	0.176	0.278	-0.0107	0.208	0.178	0.499	0.0255	0.248	0.233	0.943
0.2		-0.0602	0.198	0.175	0.135	-0.0360	0.192	0.161	0.309	-0.0082	0.196	0.159	0.537	0.0574	0.240	0.228	0.953
0.4		-0.0805	0.186	0.166	0.154	-0.0449	0.180	0.144	0.344	-0.0057	0.184	0.138	0.571	0.0982	0.225	0.221	0.963
0.6		-0.0923	0.178	0.161	0.169	-0.0499	0.173	0.136	0.363	-0.0042	0.176	0.127	0.588	0.1265	0.211	0.215	0.967
0.8		-0.1004	0.172	0.158	0.174	-0.0529	0.168	0.130	0.375	-0.0031	0.171	0.119	0.604	0.1499	0.195	0.211	0.969
1		-0.1068	0.166	0.156	0.183	-0.0551	0.165	0.127	0.386	-0.0023	0.167	0.114	0.612	0.1735	0.172	0.206	0.971
1		0	-0.0121	0.210	0.193	0.087	-0.0137	0.216	0.200	0.233	-0.0141	0.223	0.212	0.448	-0.0230	0.252	0.242
	0.1	-0.0433	0.198	0.178	0.119	-0.0290	0.195	0.171	0.288	-0.0124	0.201	0.173	0.522	0.0217	0.241	0.228	0.957
	0.2	-0.0600	0.190	0.170	0.137	-0.0366	0.186	0.157	0.323	-0.0100	0.190	0.155	0.555	0.0527	0.234	0.223	0.965
	0.4	-0.0791	0.180	0.163	0.158	-0.0445	0.175	0.142	0.356	-0.0068	0.178	0.136	0.594	0.0927	0.219	0.216	0.971
	0.6	-0.0899	0.172	0.159	0.171	-0.0490	0.169	0.134	0.376	-0.0050	0.172	0.126	0.612	0.1203	0.206	0.210	0.972
	0.8	-0.0972	0.167	0.156	0.177	-0.0517	0.164	0.129	0.389	-0.0039	0.167	0.119	0.621	0.1431	0.191	0.205	0.975
	1	-0.1029	0.162	0.154	0.180	-0.0536	0.161	0.126	0.395	-0.0031	0.164	0.114	0.627	0.1662	0.169	0.200	0.976

Table G2. Impact de α_0 et ω sur les caractéristiques opératoires pour les scénarios S5 à S8 (Distribution Weibull pour les nouvelles données et les données historiques, conflit négatif), avec σ_H^2 équivalent à 66 évènements.

Bras contrôles: Conflit négatif																		
Effet traitement		S5: Effet nul HR = 1				S6: Effet décevant HR = 0.886				S7: Effet historique HR = 0.786				S8: Effet anticipé HR = 0.55				
Paramètres		Biais	SD	RMSE	Alpha	Biais	SD	RMSE	Puissance	Biais	SD	RMSE	Puissance	Biais	SD	RMSE	Puissance	
α_0	ω																	
	0	0	0.0003	0.241	0.243	0.104	0.0008	0.240	0.243	0.224	0.0036	0.245	0.247	0.382	-0.0024	0.263	0.264	0.849
	0.1	-0.0325	0.227	0.221	0.134	-0.0217	0.221	0.209	0.277	0.0035	0.223	0.207	0.448	0.0456	0.251	0.253	0.882	
	0.2	-0.0528	0.217	0.210	0.151	-0.0323	0.209	0.190	0.308	0.0035	0.211	0.185	0.478	0.0758	0.242	0.247	0.902	
	0.4	-0.0773	0.202	0.195	0.171	-0.0444	0.194	0.169	0.345	0.0034	0.195	0.160	0.513	0.1150	0.227	0.240	0.919	
	0.6	-0.0926	0.192	0.186	0.182	-0.0515	0.185	0.156	0.362	0.0034	0.186	0.144	0.534	0.1425	0.213	0.235	0.927	
	0.8	-0.1042	0.183	0.178	0.192	-0.0564	0.178	0.146	0.375	0.0032	0.179	0.134	0.550	0.1648	0.198	0.231	0.930	
1	-0.1151	0.173	0.171	0.196	-0.0603	0.172	0.139	0.386	0.0028	0.174	0.124	0.552	0.1884	0.177	0.224	0.932		
0.3	0	0.0798	0.216	0.213	0.038	0.0755	0.217	0.213	0.115	0.0619	0.223	0.221	0.265	0.0813	0.243	0.239	0.841	
	0.1	0.0388	0.209	0.190	0.052	0.0460	0.203	0.181	0.154	0.0602	0.204	0.189	0.327	0.1239	0.227	0.239	0.880	
	0.2	0.0151	0.202	0.177	0.062	0.0302	0.194	0.163	0.180	0.0599	0.193	0.170	0.362	0.1482	0.217	0.240	0.897	
	0.4	-0.0134	0.191	0.160	0.074	0.0130	0.182	0.142	0.205	0.0538	0.181	0.150	0.398	0.1769	0.202	0.242	0.914	
	0.6	-0.0316	0.182	0.147	0.080	0.0033	0.174	0.129	0.221	0.0503	0.174	0.139	0.418	0.1947	0.190	0.244	0.920	
	0.8	-0.0452	0.174	0.137	0.086	-0.0033	0.169	0.120	0.228	0.0481	0.169	0.131	0.431	0.2078	0.180	0.245	0.928	
	1	-0.0567	0.164	0.127	0.087	-0.0082	0.164	0.113	0.236	0.0465	0.166	0.125	0.441	0.2192	0.169	0.246	0.928	
0.6	0	0.1150	0.205	0.216	0.020	0.1124	0.206	0.216	0.075	0.1169	0.212	0.224	0.206	0.1212	0.233	0.245	0.824	
	0.1	0.0739	0.201	0.192	0.030	0.0796	0.195	0.184	0.107	0.0991	0.195	0.194	0.264	0.1593	0.215	0.248	0.867	
	0.2	0.0497	0.196	0.176	0.038	0.0622	0.187	0.166	0.124	0.0904	0.186	0.177	0.296	0.1802	0.205	0.250	0.889	
	0.4	0.0194	0.186	0.156	0.045	0.0430	0.177	0.144	0.144	0.0817	0.175	0.158	0.329	0.2038	0.191	0.254	0.904	
	0.6	0.0001	0.178	0.140	0.050	0.0320	0.170	0.130	0.155	0.0768	0.169	0.147	0.348	0.2178	0.181	0.257	0.911	
	0.8	-0.0143	0.169	0.127	0.050	0.0247	0.164	0.120	0.162	0.0736	0.165	0.140	0.359	0.2277	0.173	0.258	0.917	
	1	-0.0271	0.160	0.114	0.054	0.0191	0.160	0.112	0.168	0.0713	0.161	0.134	0.368	0.2359	0.165	0.260	0.918	
1	0	0.1407	0.196	0.224	0.014	0.1395	0.198	0.225	0.056	0.1448	0.204	0.233	0.168	0.1503	0.225	0.255	0.811	
	0.1	0.1006	0.195	0.199	0.019	0.1054	0.189	0.193	0.080	0.1243	0.189	0.205	0.218	0.1849	0.207	0.259	0.853	
	0.2	0.0763	0.191	0.183	0.025	0.0871	0.183	0.175	0.095	0.1141	0.181	0.188	0.250	0.2032	0.197	0.262	0.872	
	0.4	0.0454	0.183	0.160	0.029	0.0666	0.173	0.152	0.111	0.1034	0.171	0.169	0.283	0.2232	0.184	0.266	0.891	
	0.6	0.0251	0.175	0.142	0.032	0.0546	0.167	0.137	0.121	0.0975	0.165	0.158	0.296	0.2347	0.175	0.269	0.900	
	0.8	0.0096	0.166	0.127	0.036	0.0465	0.161	0.126	0.123	0.0938	0.161	0.151	0.306	0.2428	0.168	0.270	0.904	
	1	-0.0042	0.157	0.110	0.035	0.0404	0.157	0.116	0.129	0.0913	0.158	0.145	0.314	0.2489	0.162	0.272	0.908	

Table G3. Impact de α_0 et ω sur les caractéristiques opératoires pour les scénarios S9 à S12 (Distribution Weibull pour les nouvelles données et les données historiques, conflit positif), avec σ_H^2 équivalent à 66 événements.

Bras contrôles: Conflit positif																			
Effet traitement		S9: Effet nul HR = 1				S10: Effet décevant HR = 0.886				S11: Effet historique HR = 0.786				S12: Effet anticipé HR = 0.55					
Paramètres		Biais	SD	RMSE	Alpha	Biais	SD	RMSE	Puissance	Biais	SD	RMSE	Puissance	Biais	SD	RMSE	Puissance		
184	α_0																		
	ω																		
	0	0	0.0009	0.281	0.280	0.102	-0.0003	0.280	0.284	0.201	0.0008	0.286	0.291	0.354	-0.0089	0.315	0.315	0.764	
		0.1	-0.0360	0.263	0.250	0.130	-0.0221	0.256	0.240	0.252	0.0008	0.262	0.241	0.412	0.0430	0.300	0.294	0.796	
		0.2	-0.0584	0.249	0.233	0.152	-0.0338	0.240	0.216	0.276	0.0007	0.245	0.212	0.450	0.0767	0.288	0.283	0.819	
		0.4	-0.0867	0.230	0.214	0.177	-0.0477	0.219	0.186	0.309	0.0005	0.223	0.177	0.487	0.1224	0.266	0.270	0.844	
		0.6	-0.1053	0.215	0.201	0.192	-0.0567	0.205	0.167	0.332	0.0004	0.209	0.154	0.510	0.1557	0.246	0.261	0.861	
		0.8	-0.1198	0.202	0.191	0.201	-0.0632	0.195	0.153	0.341	0.0002	0.198	0.137	0.523	0.1845	0.226	0.254	0.867	
		1	-0.1333	0.187	0.182	0.210	-0.0688	0.184	0.141	0.353	0.0001	0.187	0.123	0.535	0.2167	0.192	0.247	0.872	
		0.3	0	-0.0605	0.246	0.237	0.132	-0.0640	0.248	0.244	0.274	-0.0616	0.256	0.250	0.481	-0.0817	0.287	0.286	0.905
			0.1	-0.0885	0.227	0.216	0.169	-0.0741	0.225	0.208	0.339	-0.0489	0.234	0.206	0.524	-0.0249	0.279	0.268	0.924
			0.2	-0.1042	0.215	0.206	0.193	-0.0784	0.212	0.190	0.377	-0.0416	0.220	0.182	0.560	0.0121	0.271	0.258	0.938
			0.4	-0.1218	0.200	0.197	0.220	-0.0833	0.196	0.170	0.414	-0.0339	0.203	0.155	0.597	0.0625	0.255	0.242	0.952
			0.6	-0.1322	0.190	0.192	0.240	-0.0858	0.186	0.159	0.433	-0.0291	0.192	0.138	0.618	0.0995	0.237	0.230	0.957
			0.8	-0.1395	0.183	0.188	0.248	-0.0874	0.180	0.151	0.445	-0.0261	0.184	0.127	0.629	0.1321	0.217	0.218	0.960
			1	-0.1452	0.176	0.187	0.260	-0.0887	0.175	0.146	0.454	-0.0236	0.177	0.118	0.639	0.1696	0.182	0.203	0.963
		0.6	0	-0.0892	0.231	0.234	0.163	-0.0938	0.235	0.242	0.336	-0.0926	0.243	0.248	0.562	-0.1145	0.275	0.287	0.950
			0.1	-0.1120	0.213	0.216	0.204	-0.0978	0.214	0.209	0.408	-0.0739	0.224	0.206	0.605	-0.0586	0.271	0.269	0.961
			0.2	-0.1240	0.203	0.208	0.236	-0.0989	0.201	0.192	0.445	-0.0642	0.211	0.183	0.642	-0.0218	0.265	0.257	0.970
			0.4	-0.1374	0.189	0.200	0.270	-0.1001	0.187	0.174	0.479	-0.0532	0.195	0.157	0.676	0.0296	0.251	0.238	0.979
			0.6	-0.1450	0.181	0.197	0.289	-0.1007	0.179	0.164	0.501	-0.0472	0.186	0.141	0.692	0.0683	0.235	0.222	0.981
			0.8	-0.1501	0.175	0.195	0.296	-0.1011	0.174	0.158	0.509	-0.0428	0.179	0.130	0.700	0.1031	0.215	0.205	0.981
			1	-0.1539	0.170	0.193	0.304	-0.1013	0.170	0.154	0.519	-0.0397	0.173	0.121	0.707	0.1448	0.177	0.182	0.983
		1	0	-0.1108	0.221	0.238	0.194	-0.1162	0.226	0.246	0.400	-0.1158	0.235	0.253	0.628	-0.1391	0.268	0.293	0.971
			0.1	-0.1294	0.204	0.221	0.246	-0.1158	0.206	0.215	0.467	-0.0936	0.217	0.212	0.671	-0.0851	0.265	0.275	0.979
			0.2	-0.1389	0.194	0.213	0.278	-0.1149	0.195	0.198	0.499	-0.0820	0.205	0.189	0.703	-0.0488	0.261	0.261	0.984
			0.4	-0.1497	0.182	0.207	0.313	-0.1136	0.182	0.181	0.535	-0.0690	0.191	0.163	0.731	0.0027	0.248	0.240	0.986
			0.6	-0.1554	0.175	0.204	0.328	-0.1130	0.175	0.172	0.556	-0.0618	0.182	0.147	0.743	0.0423	0.233	0.220	0.988
		0.8	-0.1592	0.170	0.202	0.338	-0.1124	0.170	0.166	0.568	-0.0567	0.175	0.136	0.756	0.0790	0.213	0.200	0.988	
		1	-0.1621	0.166	0.200	0.346	-0.1120	0.166	0.162	0.574	-0.0528	0.169	0.127	0.758	0.1251	0.173	0.167	0.989	

Table G4. Impact de α_0 et ω sur les caractéristiques opératoires pour les scénarios S1 à S4 (Distribution Weibull pour les nouvelles données et les données historiques, bras contrôles commensurables), avec σ_H^2 équivalent à 22 évènements.

		Bras contrôles commensurables																
Paramètres	Effet traitement	S1: Effet nul HR = 1				S2: Effet décevant HR = 0.886				S3: Effet historique HR = 0.786				S4: Effet anticipé HR = 0.55				
		Biais	SD	RMSE	Alpha	Biais	SD	RMSE	Puissance	Biais	SD	RMSE	Puissance	Biais	SD	RMSE	Puissance	
185	α_0 0	ω 0	0.0024	0.263	0.262	0.100	0.0015	0.268	0.267	0.208	0.0022	0.275	0.275	0.359	-0.0075	0.304	0.298	0.794
		0.1	-0.0171	0.253	0.244	0.108	-0.0048	0.252	0.244	0.213	0.0021	0.256	0.247	0.375	0.0170	0.281	0.277	0.809
		0.2	-0.0292	0.246	0.233	0.112	-0.0111	0.245	0.230	0.220	0.0020	0.248	0.232	0.384	0.0371	0.274	0.266	0.817
		0.4	-0.0439	0.238	0.220	0.117	-0.0186	0.236	0.215	0.229	0.0016	0.239	0.215	0.396	0.0632	0.262	0.250	0.823
		0.6	-0.0527	0.232	0.212	0.118	-0.0229	0.231	0.205	0.232	0.0014	0.233	0.206	0.401	0.0803	0.253	0.239	0.826
		0.8	-0.0588	0.227	0.206	0.121	-0.0259	0.227	0.199	0.237	0.0013	0.230	0.199	0.405	0.0929	0.245	0.230	0.828
		1	-0.0630	0.224	0.201	0.122	-0.0280	0.224	0.195	0.240	0.0011	0.227	0.194	0.407	0.1034	0.237	0.223	0.829
		0.3	0	-0.0058	0.232	0.214	0.084	-0.0071	0.238	0.220	0.211	-0.0070	0.245	0.231	0.396	-0.0158	0.274	0.259
	0.1		-0.0225	0.225	0.201	0.092	-0.0120	0.227	0.203	0.215	-0.0069	0.231	0.208	0.415	0.0069	0.258	0.240	0.896
	0.2		-0.0324	0.220	0.194	0.097	-0.0166	0.221	0.194	0.225	-0.0069	0.225	0.197	0.425	0.0262	0.252	0.231	0.900
	0.4		-0.0432	0.214	0.186	0.101	-0.0217	0.215	0.183	0.233	-0.0060	0.218	0.185	0.436	0.0505	0.242	0.218	0.905
	0.6		-0.0494	0.210	0.181	0.104	-0.0245	0.211	0.178	0.237	-0.0052	0.214	0.179	0.447	0.0658	0.234	0.208	0.907
	0.8		-0.0535	0.207	0.178	0.107	-0.0262	0.209	0.174	0.242	-0.0048	0.212	0.174	0.445	0.0768	0.228	0.201	0.909
	1		-0.0564	0.205	0.176	0.109	-0.0275	0.207	0.171	0.242	-0.0046	0.210	0.171	0.452	0.0857	0.221	0.195	0.911
	0.6		0	-0.0095	0.219	0.200	0.082	-0.0108	0.225	0.206	0.222	-0.0111	0.233	0.218	0.418	-0.0199	0.261	0.248
		0.1	-0.0246	0.213	0.190	0.090	-0.0147	0.215	0.191	0.224	-0.0110	0.220	0.198	0.443	0.0022	0.248	0.230	0.926
		0.2	-0.0332	0.208	0.184	0.097	-0.0188	0.210	0.183	0.235	-0.0100	0.215	0.188	0.454	0.0206	0.242	0.220	0.931
		0.4	-0.0428	0.203	0.177	0.101	-0.0230	0.205	0.174	0.244	-0.0084	0.209	0.177	0.466	0.0440	0.233	0.208	0.934
		0.6	-0.0480	0.200	0.173	0.103	-0.0252	0.202	0.170	0.249	-0.0076	0.205	0.171	0.475	0.0587	0.226	0.199	0.937
		0.8	-0.0514	0.198	0.171	0.106	-0.0267	0.200	0.167	0.252	-0.0069	0.203	0.168	0.476	0.0691	0.220	0.192	0.938
		1	-0.0538	0.196	0.169	0.106	-0.0276	0.199	0.165	0.252	-0.0066	0.202	0.165	0.477	0.0772	0.214	0.187	0.939
		1	0	-0.0121	0.210	0.193	0.087	-0.0137	0.216	0.200	0.233	-0.0141	0.223	0.212	0.448	-0.0230	0.252	0.242
	0.1		-0.0260	0.204	0.184	0.095	-0.0167	0.207	0.186	0.236	-0.0136	0.212	0.192	0.467	-0.0015	0.241	0.225	0.944
	0.2		-0.0337	0.200	0.179	0.100	-0.0202	0.203	0.178	0.245	-0.0121	0.207	0.183	0.479	0.0166	0.235	0.216	0.947
	0.4		-0.0428	0.196	0.173	0.104	-0.0239	0.198	0.170	0.256	-0.0102	0.202	0.173	0.491	0.0390	0.227	0.203	0.949
	0.6		-0.0474	0.193	0.170	0.106	-0.0258	0.195	0.166	0.259	-0.0092	0.199	0.168	0.495	0.0533	0.220	0.194	0.953
	0.8		-0.0504	0.191	0.167	0.108	-0.0270	0.194	0.164	0.262	-0.0087	0.197	0.165	0.500	0.0633	0.214	0.188	0.952
	1		-0.0525	0.190	0.166	0.111	-0.0279	0.192	0.162	0.264	-0.0083	0.195	0.163	0.502	0.0710	0.208	0.182	0.954

Table G5. Impact de α_0 et ω sur les caractéristiques opératoires pour les scénarios S5 à S8 (Distribution Weibull pour les nouvelles données et les données historiques, conflit négatif), avec σ_H^2 équivalent à 22 évènements.

		Bras contrôles: Conflit négatif																	
Paramètres	Effet traitement	S5: Effet nul HR = 1				S6: Effet décevant HR = 0.886				S7: Effet historique HR = 0.786				S8: Effet anticipé HR = 0.55					
		Biais	SD	RMSE	Alpha	Biais	SD	RMSE	Puissance	Biais	SD	RMSE	Puissance	Biais	SD	RMSE	Puissance		
186	α_0 0	ω 0	0.0003	0.241	0.243	0.104	0.0008	0.240	0.243	0.224	0.0036	0.245	0.247	0.382	-0.0024	0.263	0.264	0.849	
			0.1	-0.0174	0.233	0.228	0.110	-0.0075	0.231	0.225	0.229	0.0040	0.236	0.227	0.405	0.0217	0.256	0.251	0.858
			0.2	-0.0283	0.227	0.219	0.117	-0.0132	0.225	0.214	0.237	0.0040	0.230	0.216	0.414	0.0394	0.250	0.242	0.864
			0.4	-0.0411	0.221	0.208	0.122	-0.0196	0.219	0.202	0.246	0.0036	0.223	0.202	0.424	0.0613	0.241	0.231	0.870
			0.6	-0.0486	0.216	0.201	0.124	-0.0233	0.214	0.195	0.254	0.0037	0.218	0.195	0.430	0.0753	0.234	0.223	0.873
			0.8	-0.0536	0.213	0.197	0.127	-0.0258	0.212	0.190	0.255	0.0038	0.216	0.190	0.432	0.0853	0.228	0.217	0.877
			1	-0.0572	0.210	0.194	0.128	-0.0272	0.210	0.187	0.259	0.0038	0.214	0.186	0.436	0.0937	0.223	0.211	0.874
			0.3	0	0.0798	0.216	0.213	0.038	0.0755	0.217	0.213	0.115	0.0619	0.223	0.221	0.265	0.0813	0.243	0.239
	0.1	0.0576			0.211	0.196	0.037	0.0662	0.211	0.198	0.120	0.0577	0.216	0.206	0.283	0.1003	0.235	0.233	0.852
	0.2	0.0457			0.208	0.186	0.041	0.0585	0.207	0.188	0.127	0.0554	0.211	0.196	0.291	0.1135	0.229	0.231	0.861
	0.4	0.0320			0.202	0.175	0.042	0.0501	0.201	0.177	0.131	0.0530	0.205	0.187	0.300	0.1289	0.221	0.228	0.865
	0.6	0.0241			0.199	0.167	0.044	0.0454	0.198	0.171	0.134	0.0518	0.202	0.182	0.308	0.1379	0.216	0.226	0.868
	0.8	0.0188			0.196	0.162	0.044	0.0425	0.196	0.167	0.137	0.0510	0.200	0.178	0.310	0.1442	0.212	0.224	0.871
	1	0.0151			0.194	0.159	0.045	0.0404	0.195	0.164	0.137	0.0500	0.199	0.174	0.312	0.1484	0.209	0.223	0.873
	0.6	0	0.1150	0.205	0.216	0.020	0.1124	0.206	0.216	0.075	0.1169	0.212	0.224	0.206	0.1212	0.233	0.245	0.824	
			0.1	0.0928	0.201	0.198	0.021	0.1017	0.201	0.202	0.079	0.1095	0.206	0.209	0.222	0.1377	0.224	0.242	0.838
			0.2	0.0809	0.198	0.188	0.022	0.0934	0.197	0.192	0.083	0.1055	0.202	0.201	0.231	0.1487	0.219	0.241	0.844
			0.4	0.0670	0.194	0.175	0.023	0.0843	0.193	0.181	0.090	0.1013	0.197	0.191	0.241	0.1611	0.212	0.239	0.849
			0.6	0.0589	0.190	0.167	0.024	0.0794	0.190	0.175	0.092	0.0991	0.194	0.187	0.244	0.1683	0.208	0.239	0.854
			0.8	0.0537	0.188	0.162	0.025	0.0762	0.189	0.171	0.092	0.0977	0.193	0.184	0.248	0.1729	0.204	0.238	0.855
			1	0.0498	0.186	0.158	0.024	0.0740	0.187	0.168	0.093	0.0968	0.191	0.182	0.252	0.1763	0.202	0.238	0.857
	1	0	0.1407	0.196	0.224	0.014	0.1395	0.198	0.225	0.056	0.1448	0.204	0.233	0.168	0.1503	0.225	0.255	0.811	
			0.1	0.1189	0.193	0.206	0.013	0.1280	0.194	0.211	0.059	0.1361	0.198	0.219	0.182	0.1650	0.217	0.253	0.822
			0.2	0.1071	0.191	0.195	0.015	0.1195	0.191	0.202	0.062	0.1315	0.195	0.211	0.190	0.1744	0.212	0.253	0.828
			0.4	0.0929	0.187	0.182	0.016	0.1101	0.187	0.191	0.065	0.1266	0.191	0.202	0.196	0.1851	0.205	0.253	0.833
			0.6	0.0848	0.184	0.174	0.016	0.1049	0.184	0.184	0.068	0.1240	0.188	0.198	0.202	0.1909	0.202	0.253	0.837
			0.8	0.0796	0.182	0.168	0.017	0.1016	0.183	0.180	0.070	0.1225	0.187	0.195	0.204	0.1949	0.199	0.253	0.840
			1	0.0756	0.183	0.164	0.017	0.0993	0.181	0.177	0.070	0.1212	0.186	0.193	0.206	0.1976	0.197	0.253	0.841

Table G6. Impact de α_0 et ω sur les caractéristiques opératoires pour les scénarios S9 à S12 (Distribution Weibull pour les nouvelles données et les données historiques, conflit positif), avec σ_H^2 équivalent à 22 évènements.

Bras contrôles: Conflit positif																	
Effet traitement		S9: Effet nul HR = 1				S10: Effet décevant HR = 0.886				S11: Effet historique HR = 0.786				S12: Effet anticipé HR = 0.55			
Paramètres		Biais	SD	RMSE	Alpha	Biais	SD	RMSE	Puissance	Biais	SD	RMSE	Puissance	Biais	SD	RMSE	Puissance
0	α_0																
	ω																
	0	0.0009	0.281	0.280	0.102	-0.0003	0.280	0.284	0.201	0.0008	0.286	0.291	0.354	-0.0089	0.315	0.315	0.764
	0.1	-0.0197	0.270	0.260	0.109	-0.0113	0.268	0.259	0.217	0.0007	0.274	0.263	0.368	0.0182	0.303	0.296	0.770
	0.2	-0.0327	0.262	0.247	0.112	-0.0180	0.259	0.244	0.223	0.0006	0.265	0.247	0.378	0.0398	0.295	0.282	0.777
	0.4	-0.0489	0.252	0.232	0.117	-0.0259	0.249	0.225	0.232	0.0006	0.254	0.226	0.387	0.0689	0.281	0.264	0.783
	0.6	-0.0586	0.245	0.222	0.121	-0.0306	0.242	0.214	0.235	0.0006	0.247	0.214	0.392	0.0881	0.270	0.251	0.788
0.8	-0.0655	0.239	0.215	0.124	-0.0338	0.238	0.206	0.235	0.0005	0.242	0.205	0.398	0.1028	0.260	0.241	0.791	
1	-0.0706	0.235	0.209	0.125	-0.0363	0.234	0.201	0.238	0.0003	0.239	0.199	0.404	0.1153	0.250	0.232	0.793	
0.3	0	-0.0605	0.246	0.237	0.132	-0.0640	0.248	0.244	0.274	-0.0616	0.256	0.250	0.481	-0.0817	0.287	0.286	0.905
	0.1	-0.0739	0.236	0.224	0.141	-0.0677	0.237	0.225	0.291	-0.0599	0.246	0.230	0.484	-0.0489	0.279	0.266	0.909
	0.2	-0.0820	0.230	0.218	0.147	-0.0697	0.231	0.215	0.299	-0.0558	0.239	0.217	0.495	-0.0256	0.272	0.251	0.914
	0.4	-0.0907	0.223	0.210	0.154	-0.0719	0.223	0.203	0.309	-0.0508	0.230	0.201	0.504	0.0051	0.261	0.230	0.918
	0.6	-0.0957	0.219	0.206	0.157	-0.0732	0.219	0.197	0.317	-0.0481	0.225	0.193	0.510	0.0259	0.251	0.214	0.920
	0.8	-0.0990	0.216	0.203	0.159	-0.0740	0.216	0.193	0.320	-0.0463	0.222	0.187	0.515	0.0416	0.242	0.200	0.921
	1	-0.1012	0.214	0.201	0.162	-0.0746	0.214	0.191	0.321	-0.0451	0.220	0.183	0.517	0.0547	0.232	0.188	0.923
0.6	0	-0.0892	0.231	0.234	0.163	-0.0938	0.235	0.242	0.336	-0.0926	0.243	0.248	0.562	-0.1145	0.275	0.287	0.950
	0.1	-0.0996	0.223	0.224	0.174	-0.0944	0.225	0.225	0.357	-0.0878	0.234	0.230	0.566	-0.0821	0.269	0.266	0.951
	0.2	-0.1053	0.218	0.218	0.182	-0.0949	0.220	0.216	0.366	-0.0823	0.228	0.217	0.573	-0.0585	0.263	0.250	0.954
	0.4	-0.1116	0.212	0.213	0.185	-0.0949	0.213	0.206	0.379	-0.0757	0.221	0.203	0.585	-0.0273	0.252	0.227	0.958
	0.6	-0.1151	0.208	0.210	0.191	-0.0951	0.209	0.201	0.385	-0.0724	0.216	0.194	0.594	-0.0061	0.243	0.208	0.958
	0.8	-0.1173	0.206	0.208	0.195	-0.0951	0.207	0.198	0.387	-0.0700	0.213	0.189	0.594	0.0102	0.234	0.192	0.958
	1	-0.1189	0.204	0.206	0.194	-0.0954	0.206	0.195	0.392	-0.0684	0.211	0.185	0.596	0.0240	0.223	0.177	0.960
1	0	-0.1108	0.221	0.238	0.194	-0.1162	0.226	0.246	0.400	-0.1158	0.235	0.253	0.628	-0.1391	0.268	0.293	0.971
	0.1	-0.1189	0.214	0.229	0.208	-0.1148	0.217	0.231	0.415	-0.1092	0.226	0.235	0.632	-0.1073	0.262	0.272	0.973
	0.2	-0.1232	0.209	0.224	0.219	-0.1139	0.212	0.222	0.425	-0.1028	0.221	0.223	0.640	-0.0843	0.257	0.256	0.973
	0.4	-0.1282	0.204	0.219	0.225	-0.1129	0.206	0.213	0.434	-0.0956	0.214	0.209	0.650	-0.0525	0.247	0.230	0.975
	0.6	-0.1308	0.201	0.217	0.231	-0.1124	0.203	0.208	0.439	-0.0914	0.210	0.200	0.657	-0.0309	0.237	0.209	0.976
	0.8	-0.1324	0.199	0.215	0.231	-0.1120	0.201	0.204	0.443	-0.0887	0.207	0.195	0.661	-0.0140	0.228	0.191	0.977
	1	-0.1334	0.197	0.214	0.232	-0.1118	0.199	0.202	0.448	-0.0868	0.205	0.191	0.665	0.0004	0.218	0.173	0.977

— Annexe H —

Impact d'un conflit
en matière de distributions de survie
entre les données historiques contrôle
et les données actuelles

Table H1. Impact de α_0 et ω sur les caractéristiques opératoires pour les scénarios S25 à S28 (Distribution Weibull pour les données historiques / exponentielle par morceaux pour les nouvelles données) avec une analyse Bayésienne Weibull (W/P/W)

Données historiques: Distribution Weibull / Nouvelles données: Distribution exponentielle par morceaux / Analyse Bayésienne Weibull (W/P/W)																		
Effet traitement		S25: Effet nul HR = 1				S26: Effet décevant HR = 0.886				S27: Effet historique HR = 0.786				S28: Effet anticipé HR = 0.55				
Paramètres		Biais	SD	RMSE	Alpha	Biais	SD	RMSE	Puissance	Biais	SD	RMSE	Puissance	Biais	SD	RMSE	Puissance	
α_0	ω																	
	0	0	-0.0015	0.277	0.287	0.113	-0.0088	0.284	0.294	0.220	-0.0142	0.292	0.298	0.358	-0.0344	0.326	0.327	0.776
	0.1		-0.0497	0.253	0.254	0.223	-0.0332	0.249	0.236	0.370	-0.0109	0.252	0.232	0.521	0.0315	0.309	0.313	0.880
	0.2		-0.0788	0.234	0.240	0.331	-0.0477	0.224	0.207	0.491	-0.0085	0.225	0.195	0.650	0.0745	0.292	0.308	0.932
	0.4		-0.1179	0.204	0.226	0.507	-0.0667	0.190	0.170	0.676	-0.0063	0.188	0.148	0.809	0.1355	0.262	0.306	0.978
	0.6		-0.1457	0.179	0.220	0.649	-0.0795	0.163	0.147	0.796	-0.0041	0.161	0.114	0.894	0.1834	0.233	0.307	0.991
	0.8		-0.1716	0.153	0.215	0.791	-0.0910	0.139	0.129	0.893	-0.0028	0.138	0.082	0.953	0.2304	0.199	0.310	0.997
	1		-0.2071	0.106	0.211	0.967	-0.1045	0.107	0.112	0.992	-0.0016	0.108	0.039	0.997	0.3125	0.112	0.315	1
0.3	0		-0.0387	0.243	0.238	0.120	-0.0472	0.250	0.248	0.262	-0.0541	0.258	0.255	0.447	-0.0784	0.292	0.294	0.893
	0.1		-0.0849	0.219	0.215	0.259	-0.0652	0.214	0.196	0.447	-0.0394	0.220	0.193	0.638	-0.0076	0.285	0.287	0.958
	0.2		-0.1108	0.201	0.208	0.377	-0.0747	0.191	0.170	0.577	-0.0311	0.195	0.159	0.760	0.0381	0.275	0.286	0.982
	0.4		-0.1429	0.175	0.204	0.544	-0.0863	0.162	0.144	0.750	-0.0214	0.164	0.119	0.888	0.1035	0.253	0.288	0.995
	0.6		-0.1648	0.153	0.204	0.694	-0.0936	0.140	0.130	0.856	-0.0153	0.142	0.092	0.942	0.1558	0.229	0.291	0.999
	0.8		-0.1830	0.133	0.205	0.816	-0.0993	0.123	0.120	0.929	-0.0106	0.124	0.066	0.977	0.2072	0.197	0.294	1
	1		-0.2042	0.103	0.208	0.966	-0.1054	0.104	0.113	0.992	-0.0061	0.104	0.040	0.998	0.2983	0.107	0.301	1
	0.6	0		-0.0563	0.229	0.228	0.138	-0.0656	0.236	0.239	0.300	-0.0736	0.245	0.248	0.505	-0.1001	0.278	0.290
0.1			-0.0994	0.205	0.209	0.286	-0.0790	0.201	0.189	0.495	-0.0530	0.208	0.187	0.700	-0.0310	0.276	0.285	0.976
0.2			-0.1230	0.188	0.203	0.407	-0.0857	0.179	0.166	0.629	-0.0420	0.185	0.156	0.809	0.0144	0.269	0.284	0.990
0.4			-0.1513	0.163	0.201	0.583	-0.0938	0.152	0.142	0.781	-0.0293	0.157	0.118	0.910	0.0797	0.252	0.284	0.998
0.6			-0.1703	0.144	0.202	0.715	-0.0985	0.133	0.129	0.873	-0.0212	0.137	0.091	0.957	0.1331	0.230	0.287	1
0.8			-0.1859	0.127	0.204	0.829	-0.1026	0.118	0.121	0.939	-0.0149	0.121	0.067	0.982	0.1878	0.200	0.289	1
1			-0.2033	0.101	0.208	0.961	-0.1067	0.102	0.115	0.991	-0.0094	0.102	0.041	0.999	0.2898	0.104	0.292	1
1		0		-0.0695	0.219	0.226	0.153	-0.0796	0.227	0.238	0.340	-0.0883	0.235	0.247	0.554	-0.1168	0.268	0.291
	0.1		-0.1092	0.195	0.208	0.310	-0.0888	0.193	0.189	0.537	-0.0633	0.201	0.188	0.742	-0.0500	0.270	0.286	0.985
	0.2		-0.1306	0.179	0.203	0.433	-0.0935	0.172	0.166	0.661	-0.0503	0.180	0.157	0.839	-0.0058	0.265	0.285	0.994
	0.4		-0.1573	0.156	0.202	0.601	-0.0995	0.146	0.144	0.804	-0.0356	0.152	0.120	0.929	0.0585	0.251	0.285	0.998
	0.6		-0.1741	0.138	0.203	0.734	-0.1027	0.129	0.131	0.889	-0.0264	0.134	0.094	0.965	0.1125	0.231	0.286	0.999
	0.8		-0.1877	0.123	0.205	0.839	-0.1051	0.115	0.123	0.945	-0.0193	0.119	0.071	0.986	0.1688	0.203	0.286	1
	1		-0.2031	0.100	0.208	0.957	-0.1081	0.101	0.116	0.990	-0.0123	0.101	0.043	0.998	0.2826	0.102	0.285	1

Table H2. Impact de α_0 et ω sur les caractéristiques opératoires pour les scénarios S25 à S28 (Distribution Weibull pour les données historiques / exponentielle par morceaux pour les nouvelles données) avec une analyse Bayésienne exponentielle par morceaux (W/P/P)

Données historiques: Distribution Weibull / Nouvelles données: Distribution exponentielle par morceaux / Analyse Bayésienne exponentielle par morceaux (W/P/P)																		
Effet traitement		S25: Effet nul HR = 1				S26: Effet décevant HR = 0.886				S27: Effet historique HR = 0.786				S28: Effet anticipé HR = 0.55				
Paramètres		Biais	SD	RMSE	Alpha	Biais	SD	RMSE	Puissance	Biais	SD	RMSE	Puissance	Biais	SD	RMSE	Puissance	
α_0	ω																	
	0	0	-0.0014	0.274	0.280	0.104	-0.0056	0.280	0.287	0.212	-0.0087	0.287	0.292	0.345	-0.0230	0.313	0.321	0.766
		0.1	-0.0510	0.249	0.248	0.217	-0.0316	0.244	0.230	0.366	-0.0065	0.247	0.226	0.516	0.0429	0.298	0.306	0.875
		0.2	-0.0810	0.230	0.235	0.327	-0.0468	0.220	0.200	0.486	-0.0047	0.220	0.188	0.643	0.0859	0.283	0.302	0.932
		0.4	-0.1202	0.200	0.222	0.508	-0.0662	0.185	0.165	0.670	-0.0034	0.183	0.142	0.810	0.1462	0.254	0.302	0.978
		0.6	-0.1481	0.175	0.216	0.653	-0.0793	0.159	0.143	0.798	-0.0021	0.155	0.109	0.897	0.1932	0.225	0.305	0.991
		0.8	-0.1732	0.148	0.213	0.794	-0.0906	0.134	0.127	0.897	-0.0009	0.131	0.078	0.953	0.2388	0.191	0.309	0.998
		1	-0.2070	0.102	0.211	0.972	-0.1040	0.102	0.111	0.992	-0.0005	0.103	0.038	0.998	0.3139	0.104	0.316	1
0.3	0	0	-0.0255	0.241	0.231	0.100	-0.0317	0.248	0.241	0.240	-0.0366	0.256	0.248	0.416	-0.0556	0.285	0.285	0.874
		0.1	-0.0750	0.218	0.209	0.234	-0.0540	0.213	0.188	0.417	-0.0262	0.216	0.185	0.618	0.0158	0.278	0.279	0.956
		0.2	-0.1028	0.200	0.202	0.348	-0.0660	0.190	0.164	0.553	-0.0203	0.191	0.152	0.740	0.0613	0.268	0.280	0.979
		0.4	-0.1375	0.174	0.199	0.529	-0.0801	0.159	0.139	0.731	-0.0137	0.160	0.113	0.874	0.1246	0.245	0.284	0.995
		0.6	-0.1604	0.153	0.200	0.676	-0.0887	0.138	0.125	0.841	-0.0092	0.138	0.085	0.938	0.1742	0.221	0.288	0.999
		0.8	-0.1800	0.132	0.202	0.809	-0.0959	0.120	0.117	0.923	-0.0061	0.120	0.063	0.976	0.2223	0.189	0.294	1
		1	-0.2020	0.101	0.206	0.966	-0.1028	0.101	0.110	0.992	-0.0030	0.101	0.038	0.998	0.3018	0.102	0.304	1
	0.6	0	0	-0.0410	0.228	0.221	0.114	-0.0483	0.235	0.232	0.268	-0.0544	0.243	0.240	0.464	-0.0766	0.273	0.281
		0.1	-0.0873	0.205	0.202	0.256	-0.0663	0.200	0.182	0.462	-0.0385	0.206	0.179	0.669	-0.0060	0.271	0.277	0.973
		0.2	-0.1133	0.188	0.198	0.376	-0.0760	0.179	0.160	0.597	-0.0304	0.183	0.148	0.785	0.0391	0.264	0.277	0.989
		0.4	-0.1442	0.164	0.197	0.554	-0.0865	0.151	0.137	0.759	-0.0204	0.153	0.111	0.902	0.1030	0.245	0.280	0.997
		0.6	-0.1649	0.145	0.198	0.696	-0.0934	0.132	0.125	0.863	-0.0147	0.134	0.085	0.951	0.1540	0.223	0.284	0.999
		0.8	-0.1818	0.126	0.201	0.816	-0.0983	0.116	0.117	0.932	-0.0097	0.117	0.063	0.979	0.2045	0.193	0.289	1
		1	-0.2006	0.100	0.205	0.960	-0.1036	0.100	0.111	0.990	-0.0058	0.100	0.040	0.998	0.2938	0.101	0.296	1
1		0	0	-0.0546	0.219	0.219	0.133	-0.0630	0.226	0.230	0.305	-0.0702	0.235	0.239	0.518	-0.0949	0.265	0.282
		0.1	-0.0979	0.196	0.202	0.282	-0.0769	0.192	0.182	0.499	-0.0497	0.199	0.180	0.715	-0.0265	0.266	0.279	0.982
		0.2	-0.1214	0.180	0.198	0.400	-0.0841	0.171	0.160	0.627	-0.0392	0.177	0.149	0.821	0.0177	0.261	0.279	0.993
		0.4	-0.1498	0.157	0.197	0.576	-0.0924	0.145	0.138	0.785	-0.0274	0.150	0.114	0.918	0.0813	0.245	0.281	0.998
		0.6	-0.1687	0.139	0.199	0.713	-0.0972	0.128	0.126	0.878	-0.0196	0.131	0.088	0.960	0.1333	0.225	0.283	1
		0.8	-0.1836	0.122	0.201	0.823	-0.1010	0.114	0.119	0.939	-0.0140	0.116	0.066	0.983	0.1862	0.197	0.286	1
		1	-0.2003	0.099	0.205	0.953	-0.1050	0.099	0.113	0.988	-0.0088	0.099	0.042	0.998	0.2867	0.100	0.289	1

Table H3. Impact de α_0 et ω sur les caractéristiques opératoires pour les scénarios S29 à S32 (Distribution exponentielle par morceaux pour les données historiques / Distribution Weibull pour les nouvelles données) avec une analyse Bayésienne par morceaux (P/W/P)

Données historiques : Distribution exponentielle par morceaux / Nouvelles données: Distribution Weibull / Analyse Bayésienne par morceaux (P/W/P)																		
Effet traitement		S29: Effet nul HR = 1				S30: Effet décevant HR = 0.886				S31: Effet historique HR = 0.786				S32: Effet anticipé HR = 0.55				
Paramètre		Biais	SD	RMSE	Alpha	Biais	SD	RMSE	Puissance	Biais	SD	RMSE	Puissance	Biais	SD	RMSE	Puissance	
α_0	ω																	
	0	0	0.0031	0.260	0.261	0.099	0	0.265	0.269	0.210	-0.0032	0.270	0.274	0.355	-0.0161	0.292	0.295	0.798
	0.1	-0.0493	0.237	0.234	0.203	-0.0282	0.231	0.215	0.373	-0.0021	0.232	0.209	0.550	0.0487	0.280	0.286	0.903	
	0.2	-0.0796	0.220	0.224	0.315	-0.0440	0.208	0.188	0.503	-0.0016	0.207	0.174	0.667	0.0904	0.267	0.284	0.951	
	0.4	-0.1189	0.192	0.215	0.501	-0.0639	0.176	0.157	0.680	-0.0012	0.173	0.131	0.805	0.1486	0.242	0.287	0.983	
	0.6	-0.1465	0.168	0.211	0.653	-0.0776	0.152	0.138	0.798	-0.0010	0.148	0.101	0.887	0.1938	0.216	0.292	0.994	
	0.8	-0.1707	0.145	0.209	0.786	-0.0889	0.130	0.123	0.887	-0.0010	0.127	0.074	0.955	0.2378	0.185	0.298	0.999	
	1	-0.2033	0.101	0.207	0.966	-0.1020	0.102	0.110	0.990	0	0.102	0.040	0.998	0.3090	0.103	0.311	1	
0.3	0	0.0040	0.230	0.210	0.070	0.0020	0.236	0.217	0.189	0.0005	0.242	0.224	0.378	-0.0051	0.267	0.250	0.873	
	0.1	-0.0525	0.212	0.193	0.186	-0.0299	0.203	0.169	0.381	0.0012	0.203	0.163	0.591	0.0645	0.258	0.252	0.958	
	0.2	-0.0838	0.197	0.189	0.292	-0.0466	0.182	0.148	0.516	0.0014	0.180	0.132	0.711	0.1072	0.246	0.257	0.982	
	0.4	-0.1222	0.173	0.188	0.482	-0.0659	0.154	0.127	0.700	0.0016	0.150	0.097	0.852	0.1643	0.223	0.268	0.995	
	0.6	-0.1478	0.153	0.190	0.638	-0.0780	0.135	0.115	0.823	0.0019	0.131	0.074	0.925	0.2068	0.199	0.278	0.999	
	0.8	-0.1693	0.132	0.193	0.777	-0.0873	0.119	0.109	0.908	0.0022	0.116	0.056	0.971	0.2461	0.170	0.288	1	
	1	-0.1940	0.100	0.198	0.951	-0.0961	0.100	0.104	0.988	0.0027	0.101	0.039	0.998	0.3030	0.101	0.305	1	
0.6	0	0.0081	0.217	0.195	0.062	0.0075	0.224	0.201	0.180	0.0076	0.231	0.209	0.376	0.0056	0.256	0.236	0.894	
	0.1	-0.0488	0.202	0.181	0.170	-0.0264	0.192	0.157	0.378	0.0066	0.192	0.150	0.599	0.0754	0.247	0.243	0.969	
	0.2	-0.0801	0.189	0.179	0.280	-0.0437	0.173	0.138	0.518	0.0059	0.170	0.122	0.729	0.1173	0.236	0.251	0.987	
	0.4	-0.1182	0.167	0.181	0.464	-0.0632	0.147	0.119	0.701	0.0052	0.143	0.089	0.862	0.1723	0.213	0.264	0.998	
	0.6	-0.1437	0.148	0.184	0.618	-0.0750	0.130	0.110	0.820	0.0050	0.125	0.069	0.927	0.2126	0.190	0.275	0.999	
	0.8	-0.1648	0.129	0.187	0.766	-0.0840	0.115	0.104	0.906	0.0048	0.112	0.053	0.969	0.2495	0.163	0.286	1	
	1	-0.1885	0.099	0.193	0.940	-0.0922	0.099	0.100	0.983	0.0049	0.100	0.039	0.998	0.3009	0.100	0.303	1	
1	0	0.0131	0.208	0.187	0.057	0.0138	0.215	0.192	0.174	0.0153	0.222	0.201	0.376	0.0158	0.249	0.230	0.907	
	0.1	-0.0438	0.196	0.175	0.161	-0.0218	0.185	0.151	0.372	0.0123	0.184	0.145	0.607	0.0849	0.240	0.240	0.972	
	0.2	-0.0748	0.184	0.174	0.266	-0.0397	0.167	0.133	0.513	0.0106	0.163	0.117	0.731	0.1259	0.228	0.249	0.989	
	0.4	-0.1131	0.163	0.176	0.450	-0.0595	0.143	0.115	0.696	0.0090	0.138	0.086	0.861	0.1789	0.206	0.264	0.998	
	0.6	-0.1387	0.146	0.179	0.595	-0.0717	0.127	0.107	0.815	0.0081	0.122	0.067	0.927	0.2174	0.184	0.275	0.999	
	0.8	-0.1598	0.128	0.183	0.741	-0.0806	0.114	0.101	0.900	0.0075	0.110	0.053	0.970	0.2521	0.158	0.285	1	
	1	-0.1840	0.099	0.188	0.922	-0.0889	0.099	0.098	0.979	0.0072	0.099	0.041	0.997	0.2999	0.100	0.302	1	

Table H4. Impact de α_0 et ω sur les caractéristiques opératoires pour les scénarios S29 à S32 (Distribution exponentielle par morceaux pour les données historiques / Distribution Weibull pour les nouvelles données) avec une analyse Bayésienne Weibull (P/W/W)

Données historiques : Distribution exponentielle par morceaux / Nouvelles données: Distribution Weibull / Analyse Bayésienne Weibull (P/W/W)																		
Effet traitement		S29: Effet nul HR = 1				S30: Effet décevant HR = 0.886				S31: Effet historique HR = 0.786				S32: Effet anticipé HR = 0.55				
Paramètres		Biais	SD	RMSE	Alpha	Biais	SD	RMSE	Puissance	Biais	SD	RMSE	Puissance	Biais	SD	RMSE	Puissance	
α_0	ω																	
	0	0	0.0032	0.262	0.257	0.098	0.0009	0.268	0.266	0.208	0.0004	0.275	0.276	0.347	-0.0084	0.275	0.294	0.793
	0.1	-0.0498	0.239	0.231	0.204	-0.0273	0.233	0.212	0.376	-0.0010	0.235	0.211	0.529	0.0548	0.288	0.284	0.898	
	0.2	-0.0807	0.222	0.221	0.313	-0.0436	0.210	0.185	0.500	-0.0007	0.210	0.176	0.658	0.0966	0.272	0.284	0.949	
	0.4	-0.1199	0.194	0.212	0.499	-0.0637	0.178	0.154	0.684	-0.0007	0.176	0.133	0.806	0.1548	0.244	0.287	0.983	
	0.6	-0.1476	0.171	0.209	0.652	-0.0776	0.155	0.135	0.798	-0.0006	0.151	0.102	0.898	0.1996	0.218	0.293	0.995	
	0.8	-0.1714	0.147	0.208	0.788	-0.0890	0.133	0.122	0.889	-0.0003	0.131	0.074	0.959	0.2422	0.186	0.299	0.999	
	1	-0.2033	0.105	0.207	0.968	-0.1018	0.106	0.109	0.989	0	0.107	0.039	0.998	0.3099	0.110	0.312	1	
0.3	0	0.0065	0.232	0.208	0.067	0.0054	0.238	0.215	0.183	0.0068	0.245	0.227	0.354	0	0.273	0.249	0.870	
	0.1	-0.0505	0.214	0.191	0.181	-0.0275	0.205	0.168	0.375	0.0047	0.205	0.166	0.572	0.0680	0.261	0.252	0.955	
	0.2	-0.0823	0.199	0.188	0.287	-0.0446	0.184	0.147	0.509	0.0041	0.182	0.136	0.705	0.1109	0.248	0.258	0.980	
	0.4	-0.1210	0.175	0.187	0.477	-0.0647	0.156	0.126	0.695	0.0037	0.153	0.100	0.851	0.1678	0.223	0.269	0.995	
	0.6	-0.1469	0.155	0.189	0.637	-0.0770	0.137	0.115	0.818	0.0034	0.133	0.076	0.927	0.2100	0.198	0.279	0.999	
	0.8	-0.1687	0.135	0.192	0.776	-0.0865	0.121	0.108	0.907	0.0034	0.118	0.057	0.968	0.2486	0.169	0.289	1	
	1	-0.1934	0.102	0.197	0.951	-0.0954	0.103	0.103	0.987	0.0033	0.103	0.039	0.998	0.3036	0.105	0.306	1	
	0.6	0	0.0106	0.219	0.194	0.063	0.0102	0.225	0.201	0.176	0.0124	0.232	0.213	0.362	0.0074	0.260	0.237	0.894
0.1		-0.0465	0.204	0.181	0.171	-0.0245	0.193	0.157	0.372	0.0093	0.193	0.155	0.589	0.0758	0.249	0.244	0.969	
0.2		-0.0783	0.190	0.179	0.275	-0.0421	0.174	0.138	0.514	0.0081	0.171	0.126	0.717	0.1176	0.237	0.252	0.987	
0.4		-0.1168	0.169	0.180	0.463	-0.0621	0.149	0.119	0.697	0.0067	0.145	0.093	0.861	0.1727	0.214	0.266	0.998	
0.6		-0.1425	0.150	0.183	0.612	-0.0742	0.132	0.109	0.821	0.0062	0.127	0.072	0.928	0.2131	0.190	0.276	0.999	
0.8		-0.1639	0.131	0.187	0.762	-0.0833	0.117	0.103	0.904	0.0058	0.114	0.055	0.971	0.2496	0.163	0.286	1	
1		-0.1880	0.101	0.192	0.939	-0.0917	0.101	0.100	0.983	0.0058	0.102	0.040	0.995	0.3008	0.103	0.303	1	
1		0	0.0145	0.209	0.188	0.058	0.0148	0.216	0.193	0.176	0.0175	0.223	0.207	0.369	0.0138	0.251	0.231	0.909
	0.1	-0.0421	0.197	0.176	0.160	-0.0210	0.186	0.151	0.370	0.0133	0.186	0.149	0.594	0.0815	0.241	0.241	0.973	
	0.2	-0.0734	0.185	0.174	0.267	-0.0389	0.168	0.133	0.514	0.0114	0.165	0.122	0.724	0.1224	0.230	0.250	0.989	
	0.4	-0.1118	0.165	0.176	0.445	-0.0590	0.145	0.116	0.692	0.0095	0.140	0.090	0.862	0.1757	0.208	0.264	0.998	
	0.6	-0.1375	0.148	0.179	0.593	-0.0712	0.129	0.107	0.813	0.0085	0.124	0.071	0.926	0.2145	0.185	0.275	0.999	
	0.8	-0.1590	0.129	0.183	0.739	-0.0803	0.115	0.101	0.899	0.0078	0.112	0.055	0.968	0.2498	0.159	0.285	1	
	1	-0.1836	0.100	0.188	0.920	-0.0887	0.100	0.097	0.978	0.0076	0.100	0.041	0.995	0.2988	0.101	0.301	1	

— Annexe I —

Impact des données historiques
contrôles réelles
scénarios S33-S44

Table I1. Impact de α_0 et ω sur les caractéristiques opératoires pour les scénarios S33 à S36 (Bras contrôles commensurables).

		Bras contrôles commensurables															
Paramètres	Effet traitement	S33: Effet Nul HR = 1				S34: Effet décevant HR = 0.886				S35: Effet historique HR = 0.786				S36: Effet espéré HR = 0.55			
	α_0	ω	Biais	SD	RMSE	Alpha	Biais	SD	RMSE	Puissance	Biais	SD	RMSE	Puissance	Biais	SD	RMSE
0	0	0.0002	0.246	0.251	0.104	-0.0068	0.251	0.255	0.227	-0.0019	0.257	0.259	0.381	-0.0148	0.279	0.283	0.824
	0.05	-0.0298	0.235	0.236	0.161	-0.0234	0.233	0.223	0.313	-0.0010	0.236	0.221	0.484	0.0226	0.274	0.279	0.885
	0.1	-0.0503	0.225	0.227	0.222	-0.0345	0.278	0.203	0.385	-0.0011	0.219	0.196	0.569	0.0504	0.268	0.278	0.922
	0.15	-0.0666	0.216	0.222	0.279	-0.0431	0.206	0.189	0.453	-0.0006	0.206	0.178	0.633	0.0727	0.262	0.279	0.948
0.1	0	-0.0046	0.229	0.220	0.090	-0.0110	0.235	0.226	0.224	-0.0058	0.241	0.230	0.397	-0.0188	0.265	0.257	0.882
	0.05	-0.0360	0.219	0.208	0.152	-0.0283	0.216	0.196	0.320	-0.0046	0.219	0.193	0.513	0.0206	0.261	0.257	0.929
	0.1	-0.0575	0.210	0.202	0.214	-0.0396	0.202	0.178	0.401	-0.0039	0.203	0.169	0.600	0.0494	0.257	0.259	0.958
	0.15	-0.0737	0.202	0.199	0.272	-0.0478	0.191	0.166	0.474	-0.0029	0.190	0.152	0.666	0.0725	0.251	0.261	0.971
0.2	0	-0.0069	0.219	0.205	0.089	-0.0125	0.225	0.212	0.221	-0.0077	0.231	0.216	0.411	-0.0202	0.256	0.245	0.906
	0.05	-0.0385	0.210	0.195	0.146	-0.0299	0.207	0.184	0.319	-0.0058	0.209	0.180	0.537	0.0199	0.254	0.246	0.950
	0.1	-0.0602	0.201	0.190	0.212	-0.0415	0.193	0.167	0.412	-0.0046	0.193	0.157	0.624	0.0487	0.250	0.249	0.971
	0.15	-0.0762	0.194	0.188	0.265	-0.0496	0.182	0.155	0.482	-0.0037	0.181	0.140	0.689	0.0716	0.245	0.252	0.982
0.3	0	-0.0076	0.212	0.196	0.086	-0.0131	0.218	0.204	0.227	-0.0084	0.225	0.208	0.424	-0.0209	0.250	0.238	0.921
	0.05	-0.0396	0.203	0.187	0.147	-0.0308	0.200	0.177	0.332	-0.0063	0.203	0.172	0.551	0.0195	0.249	0.241	0.961
	0.1	-0.0610	0.195	0.184	0.210	-0.0421	0.187	0.160	0.421	-0.0049	0.187	0.150	0.638	0.0482	0.245	0.244	0.978
	0.15	-0.0771	0.188	0.183	0.264	-0.0503	0.176	0.150	0.493	-0.0040	0.175	0.134	0.710	0.0709	0.241	0.247	0.987
0.4	0	-0.0080	0.207	0.190	0.086	-0.0135	0.213	0.200	0.229	-0.0085	0.220	0.203	0.434	-0.0210	0.246	0.233	0.934
	0.05	-0.0399	0.199	0.183	0.151	-0.0309	0.196	0.173	0.335	-0.0062	0.198	0.168	0.560	0.0190	0.245	0.237	0.968
	0.1	-0.0613	0.191	0.180	0.208	-0.0424	0.182	0.157	0.428	-0.0053	0.183	0.146	0.650	0.0478	0.242	0.241	0.983
	0.15	-0.0772	0.184	0.179	0.261	-0.0502	0.172	0.146	0.500	-0.0037	0.171	0.130	0.719	0.0702	0.237	0.245	0.989

Table I2. Impact de α_0 et ω sur les caractéristiques opératoires pour les scénarios S37 à S40 (Conflit négatif).

Bras contrôles : Conflit négatif																	
Effet traitement Paramètres		S37: Effet Nul HR = 1				S38: Effet décevant HR = 0.886				S39: Effet historique HR = 0.786				S40: Effet espéré HR = 0.55			
		Biais	SD	RMSE	Alpha	Biais	SD	RMSE	Puissance	Biais	SD	RMSE	Puissance	Biais	SD	RMSE	Puissance
α_0 0	ω 0	0.0035	0.235	0.236	0.097	-0.0043	0.239	0.245	0.232	-0.0002	0.244	0.250	0.382	-0.0122	0.265	0.263	0.860
	0.05	-0.0277	0.225	0.223	0.157	-0.0211	0.221	0.216	0.323	-0.0002	0.224	0.213	0.495	0.0255	0.261	0.262	0.910
	0.1	-0.0491	0.216	0.216	0.215	-0.0322	0.207	0.197	0.404	-0.0006	0.208	0.189	0.577	0.0529	0.256	0.263	0.937
	0.15	-0.0656	0.208	0.212	0.275	-0.0411	0.196	0.184	0.476	-0.0006	0.196	0.171	0.637	0.0752	0.250	0.266	0.957
0.1	0	0.0451	0.220	0.213	0.057	0.0383	0.225	0.221	0.162	0.0414	0.230	0.228	0.323	0.0282	0.252	0.242	0.866
	0.05	0.0105	0.215	0.202	0.101	0.0142	0.209	0.193	0.254	0.0346	0.210	0.192	0.439	0.0665	0.246	0.245	0.917
	0.1	-0.0135	0.209	0.197	0.148	-0.0015	0.197	0.177	0.333	0.0300	0.195	0.169	0.535	0.0944	0.240	0.249	0.951
	0.15	-0.0320	0.203	0.193	0.197	-0.0130	0.187	0.165	0.409	0.0268	0.183	0.153	0.600	0.1155	0.233	0.254	0.970
0.2	0	0.0695	0.210	0.207	0.040	0.0631	0.216	0.213	0.131	0.0657	0.221	0.222	0.292	0.0518	0.243	0.235	0.869
	0.05	0.0343	0.208	0.197	0.072	0.0358	0.203	0.187	0.214	0.0544	0.202	0.186	0.406	0.0903	0.237	0.241	0.922
	0.1	0.0097	0.204	0.191	0.115	0.0179	0.192	0.171	0.290	0.0476	0.187	0.164	0.502	0.1173	0.230	0.247	0.954
	0.15	-0.0094	0.200	0.188	0.155	0.0046	0.183	0.160	0.357	0.0426	0.176	0.148	0.571	0.1379	0.223	0.252	0.973
0.3	0	0.0857	0.204	0.206	0.030	0.0798	0.209	0.212	0.112	0.0818	0.215	0.220	0.273	0.0675	0.238	0.233	0.867
	0.05	0.0512	0.204	0.196	0.057	0.0505	0.198	0.186	0.187	0.0680	0.196	0.185	0.384	0.1057	0.231	0.240	0.926
	0.1	0.0264	0.201	0.190	0.094	0.0315	0.189	0.171	0.262	0.0593	0.183	0.164	0.471	0.1319	0.223	0.247	0.958
	0.15	0.0074	0.198	0.187	0.129	0.0174	0.181	0.159	0.325	0.0532	0.172	0.148	0.549	0.1515	0.216	0.253	0.973
0.4	0	0.0975	0.199	0.207	0.024	0.0915	0.204	0.212	0.101	0.0932	0.211	0.220	0.259	0.0785	0.234	0.233	0.868
	0.05	0.0633	0.200	0.197	0.051	0.0611	0.195	0.187	0.171	0.0776	0.193	0.186	0.369	0.1165	0.226	0.241	0.930
	0.1	0.0390	0.199	0.191	0.081	0.0416	0.186	0.171	0.240	0.0679	0.179	0.165	0.460	0.1416	0.218	0.248	0.957
	0.15	0.0200	0.197	0.187	0.112	0.0270	0.179	0.160	0.306	0.0609	0.169	0.150	0.528	0.1612	0.211	0.255	0.974

Table I3. Impact de α_0 et ω sur les caractéristiques opératoires pour les scénarios S41 à S44 (Conflit positif).

Bras contrôles : Conflit positif																	
Effet traitement		S41: Effet Nul HR = 1				S42: Effet décevant HR = 0.886				S43: Effet historique HR = 0.786				S44: Effet espéré HR = 0.55			
Paramètres		Biais	SD	RMSE	Alpha	Biais	SD	RMSE	Puissance	Biais	SD	RMSE	Puissance	Biais	SD	RMSE	Puissance
α_0	ω																
0	0	-0.0042	0.261	0.262	0.100	0.0084	0.267	0.270	0.195	-0.0009	0.274	0.276	0.350	-0.0157	0.298	0.299	0.800
	0.05	-0.0343	0.248	0.244	0.164	-0.0092	0.248	0.237	0.284	-0.0006	0.252	0.236	0.444	0.0229	0.292	0.293	0.859
	0.1	-0.0555	0.237	0.235	0.218	-0.0222	0.233	0.216	0.359	-0.0001	0.235	0.209	0.531	0.0514	0.284	0.290	0.899
	0.15	-0.0723	0.227	0.229	0.275	-0.0311	0.221	0.200	0.424	-0.0007	0.221	0.188	0.598	0.0746	0.277	0.290	0.925
0.1	0	-0.0654	0.242	0.236	0.138	-0.0536	0.248	0.242	0.273	-0.0612	0.255	0.252	0.458	-0.0730	0.282	0.280	0.899
	0.05	-0.0913	0.226	0.222	0.213	-0.0642	0.226	0.210	0.377	-0.0511	0.233	0.213	0.566	-0.0321	0.281	0.275	0.939
	0.1	-0.1090	0.213	0.215	0.283	-0.0710	0.210	0.190	0.463	-0.0445	0.216	0.187	0.653	-0.0008	0.278	0.274	0.963
	0.15	-0.1222	0.202	0.212	0.352	-0.0764	0.197	0.177	0.533	-0.0395	0.203	0.169	0.721	0.0240	0.274	0.274	0.975
0.2	0	-0.0987	0.230	0.232	0.169	-0.0880	0.237	0.237	0.331	-0.0943	0.244	0.248	0.536	-0.1042	0.272	0.277	0.940
	0.05	-0.1216	0.213	0.220	0.259	-0.0929	0.215	0.206	0.444	-0.0785	0.224	0.210	0.649	-0.0634	0.275	0.273	0.969
	0.1	-0.1364	0.199	0.214	0.343	-0.0965	0.198	0.187	0.536	-0.0676	0.208	0.185	0.734	-0.0328	0.274	0.271	0.982
	0.15	-0.1473	0.188	0.212	0.415	-0.0985	0.186	0.175	0.608	-0.0604	0.195	0.168	0.796	-0.0072	0.271	0.271	0.990
0.3	0	-0.1198	0.223	0.234	0.195	-0.1090	0.229	0.238	0.379	-0.1148	0.237	0.249	0.592	-0.1234	0.266	0.278	0.962
	0.05	-0.1398	0.204	0.222	0.294	-0.1104	0.207	0.207	0.497	-0.0953	0.218	0.212	0.707	-0.0841	0.270	0.274	0.980
	0.1	-0.1526	0.191	0.217	0.380	-0.1114	0.191	0.189	0.589	-0.0824	0.203	0.188	0.784	-0.0542	0.271	0.272	0.990
	0.15	-0.1620	0.180	0.214	0.466	-0.1118	0.179	0.177	0.664	-0.0732	0.191	0.170	0.838	-0.0296	0.270	0.271	0.993
0.4	0	-0.1344	0.217	0.237	0.216	-0.1238	0.224	0.240	0.412	-0.1287	0.232	0.251	0.632	-0.1366	0.261	0.281	0.972
	0.05	-0.1521	0.199	0.225	0.321	-0.1225	0.202	0.210	0.536	-0.1070	0.214	0.215	0.744	-0.0985	0.267	0.277	0.988
	0.1	-0.1636	0.185	0.220	0.417	-0.1214	0.186	0.192	0.633	-0.0926	0.199	0.191	0.823	-0.0695	0.269	0.274	0.993
	0.15	-0.1717	0.174	0.218	0.504	-0.1209	0.175	0.180	0.698	-0.0824	0.188	0.174	0.868	-0.0449	0.268	0.273	0.996

Titre : Approche Bayésienne de la survie dans les essais cliniques pour les cancers rares

Mots clés : Bayésien, Survie, Essais cliniques, Maladies Rares, Distribution a priori, Données historiques

Résumé : L'approche Bayésienne permet d'enrichir l'information apportée par l'essai clinique, en intégrant des informations externes à l'essai. De plus, elle permet d'exprimer les résultats directement en termes de probabilité d'un certain effet du traitement, plus informative et interprétable qu'une p-valeur et un intervalle de confiance. Par ailleurs, la réduction fréquente d'une analyse à une interprétation binaire des résultats (significatif ou non) est particulièrement dommageable dans les maladies rares.

L'objectif de mon travail était d'explorer la faisabilité, les contraintes et l'apport de l'approche Bayésienne dans les essais cliniques portant sur des cancers rares lorsque le critère principal est censuré.

Tout d'abord, une revue de la littérature a confirmé la faible implémentation actuelle des méthodes Bayésiennes dans l'analyse des essais cliniques avec

critère de survie. Le second axe de ce travail a porté sur le développement d'un essai Bayésien avec critère de survie, intégrant des données historiques, dans le cadre d'un essai réel portant sur une pathologie rare (ostéosarcome). Le prior intégrait des données historiques individuelles sur le bras contrôle et des données agrégées sur l'effet relatif du traitement. Une large étude de simulations a permis d'évaluer les caractéristiques opératoires du design proposé, de calibrer le modèle, tout en explorant la problématique de la commensurabilité entre les données historiques et actuelles.

Enfin, la ré-analyse de trois essais cliniques publiés a permis d'illustrer l'apport de l'approche Bayésienne dans l'expression des résultats et la manière dont cette approche permet d'enrichir l'analyse fréquentiste d'un essai.

Title : Bayesian approach to survival in clinical trials in rare cancers

Keywords : Bayesian, Survival, Clinical trials, Rare diseases, Prior distribution, Historical data

Abstract : Bayesian approach augments the information provided by the trial itself by incorporating external information into the trial analysis. In addition, this approach allows the results to be expressed in terms of probability of some treatment effect, which is more informative and interpretable than a p-value and a confidence interval. In addition, the frequent reduction of an analysis to a binary interpretation of the results (significant versus non-significant) is particularly harmful in rare diseases. In this context, the objective of my work was to explore the feasibility, constraints and contribution of the Bayesian approach in clinical trials in rare cancers with a primary censored endpoint.

A review of the literature confirmed that the implementation of Bayesian methods is still limited in the

analysis of clinical trials with a censored endpoint. In the second part of our work, we developed a Bayesian design, integrating historical data in the setting of a real clinical trial with a survival endpoint in a rare disease (osteosarcoma). The prior incorporated individual historical data on the control arm and aggregate historical data on the relative treatment effect. Through a large simulation study, we evaluated the operating characteristics of the proposed design and calibrated the model while exploring the issue of commensurability between historical and current data.

Finally, the re-analysis of three clinical trials allowed us to illustrate the contribution of Bayesian approach to the expression of the results, and how this approach enriches the frequentist analysis of a trial.

