



**HAL**  
open science

# Développement de méthodes bayésiennes de sélection de variables pour intégrer l'information experte

Sandrine Boulet

► **To cite this version:**

Sandrine Boulet. Développement de méthodes bayésiennes de sélection de variables pour intégrer l'information experte. Statistiques [math.ST]. Université Paris Cité, 2019. Français. NNT : 2019UNIP5147 . tel-03650852

**HAL Id: tel-03650852**

**<https://theses.hal.science/tel-03650852>**

Submitted on 25 Apr 2022

**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.



UNIVERSITÉ  
PARIS  
DESCARTES

**U-S-PC**  
Université Sorbonne  
Paris Cité

UNIVERSITÉ PARIS DESCARTES

École doctorale Pierre Louis de santé publique  
ED393 Épidémiologie et Sciences de l'Information Biomédicale

*INSERM UMRS 1138 Équipe 22*  
*"Sciences de l'Information au service de la médecine personnalisée"*

# Développement de méthodes Bayésiennes de sélection de variables pour intégrer l'information experte

Par Sandrine Boulet

Thèse de doctorat de Biostatistiques

Dirigée par Sarah Zohar et Anne-Sophie Jannot

Présentée et soutenue publiquement le Jeudi 12 Septembre 2019

Devant un jury composé de :

Frédéric BERTRAND	MCU à Université de Strasbourg (Rapporteur)
Pierre PUDLO	PU à Aix-Marseille Université (Rapporteur)
Adeline LECLERCQ SAMSON	PU à Université Grenoble Alpes (Examinateur)
Erik-André SAULEAU	PU-PH à Université de Strasbourg (Examinateur)
Sarah ZOHAR	DR INSERM (Directrice de thèse)
Anne-Sophie JANNOT	MCU-PH à Université Paris-Descartes (Co-directrice de thèse)



# Remerciements

Je remercie Frédéric Bertrand et Pierre Pudlo d'avoir accepté d'évaluer ce travail et d'être rapporteurs de cette thèse ; et Adeline Leclercq-Samson et Erik-André Sauleau d'avoir accepté d'examiner ce travail. Je remercie Sarah Zohar et Anne-Sophie Jannot d'avoir co-dirigé mon travail. Elles ont passé beaucoup de temps sur mes résultats intermédiaires, ainsi qu'à revoir mes présentations et à relire nos articles. Elles étaient aussi à mes côtés dans les moments de remise en question, aussi bien professionnelle que personnelle. Elles sont encore là pour me guider et me conseiller dans mon projet d'après-thèse. Je remercie également Moreno Ursino, ancien post-doctorant de notre équipe, pour s'être tant impliqué dans la modélisation statistique et dans l'encadrement de ma thèse, mais aussi pour sa gentillesse et ses conseils.

Je remercie Peter Thall, biostatisticien au MD Anderson Cancer Center à Houston (Texas), pour nous avoir accompagnés durant tout ce projet et pour nous avoir fourni de nombreuses idées et pistes de réflexion.

Je remercie également le Dr Aziz Zannan ainsi que ses collègues, les docteurs Julien Taieb, Bruno Landi, Céline Lepère, Simon Pernot, Anne-Laure Pointet et Jean-Nicolas Vaillant, tous oncologues à l'hôpital européen Georges Pompidou à Paris. Ils m'ont apporté leur expertise clinique sur les chimiothérapies, et m'ont en particulier fourni les poids de pertinence clinique, cruciales dans ce projet. Je remercie Yann Orven, pharmacien, pour m'avoir appris à utiliser DxCare. Je remercie Céline Narjoz, biologiste, pour m'avoir expliqué les caractéristiques des traits génotypiques.

Je remercie Anita Burgun pour m'avoir accueillie dans l'équipe 22 de l'unité INSERM 1138, Jérémie Guedj pour ses conseils en tant que tuteur et Dominique Costagliola et Pierre-Yves Boëlle, directeurs successifs de l'Ecole Doctorale 393. Ma gratitude va également à l'Université Paris Descartes qui m'a fourni un contrat doctoral sur trois ans, et à l'Université Versailles/St-Quentin pour le poste ATER de ma quatrième année.

Je remercie mon ex-collègue Rihab Khemiri pour son aide en informatique, mais également pour son amitié et son dévouement. J'ai également une pensée particulière pour mes collègues Abdel-Ali Boussadi, Hector Countouris, William Digan, Brahim Hadji, Samir Mellikeche, Bastien Rance et Eric Zapletal qui m'ont chacun leur tour accordé (entre autres) leur aide pour des problèmes informatiques allant d'un mauvais branchement à l'installation d'un serveur. Je remercie Catherine Schramm pour son aide sur les modèles mixtes. Je remercie aussi Camille Aupiais, Antoine Barbieri, Alois Castro, Anaïs Charles-Nelson, Xiaoyi Chen, Gabrielle Clain, Marie De Antonio, Jean-Baptiste Escudie, Emma Gérard, Angelika Geroldinger, Phan Hoang, Rodolphe Jantzen, Sandrine Katsahian, Christina Khnaïsser, Andrea Lazzati, Vincent Looten, Antoine Neuraz, Adrien Ollier, Camille Petri, Marie-Karelle Rivière, Jason Schourick, Marie Simon, Rosy Tsopra, Raphaël Veil et Maxime Wack. Je remercie chacun de mes autres collègues pour leur bonne humeur et leur écoute.

Pour finir, merci à Manu qui m'a supportée et s'est adapté pour que ma thèse se déroule dans les meilleurs conditions. Merci aussi à Alex, Apolline, Mady, Marine, Dorian, Flavie, Mélanie, Margaux, Mehdi, Solène, Yassine et tous les autres. Merci à mes parents.



# Table des matières

<b>Tableaux</b>	<b>1</b>
<b>Graphiques</b>	<b>1</b>
<b>Production scientifique</b>	<b>1</b>
<b>Glossaire</b>	<b>3</b>
<b>1 Contexte et enjeux</b>	<b>5</b>
1.1 Quels sont les enjeux actuels des méthodes de sélection de variables pour les données de soin? . . . . .	5
1.2 Comment modéliser la décision médicale à partir d'informations expertes? . . . . .	6
1.3 Intérêt de la modélisation des décisions médicales en chimiothérapie . . . . .	7
1.4 Problématique de la thèse . . . . .	8
<b>2 État des connaissances</b>	<b>11</b>
2.1 Méthodes de sélection de variables . . . . .	11
2.1.1 Inférence fréquentiste . . . . .	12
2.1.2 Inférence Bayésienne . . . . .	14
2.2 Comment représenter l'information experte? Principe de l'élicitation. . . . .	17
2.3 Comment introduire l'information experte dans les modèles utilisant l'inférence Bayésienne? . . . . .	18
2.3.1 Introduction dans les distributions a priori . . . . .	18
2.3.2 <i>Power Prior</i> . . . . .	19
2.4 Combinaison de plusieurs expertises . . . . .	20
2.4.1 La méthode Delphi . . . . .	21
2.4.2 Approche mathématique . . . . .	21
<b>3 Présentation du cas d'usage</b>	<b>23</b>
3.1 Le cancer colorectal . . . . .	23
3.1.1 Généralités . . . . .	23
3.1.2 Facteurs de risque . . . . .	23
3.1.3 Traitements . . . . .	24
3.2 Protocoles de chimiothérapie . . . . .	24
3.2.1 Protocoles incluant de l'irinotécan . . . . .	25
3.2.2 Adaptation de la posologie . . . . .	25
3.2.3 Toxicités . . . . .	25
3.3 Données disponibles . . . . .	26
3.3.1 Patients inclus . . . . .	26

3.3.2	Variables extraites . . . . .	26
3.3.3	Analyse descriptive . . . . .	27
3.4	Élicitation des poids de sévérité . . . . .	29
4	Méthode de sélection de variables Bayésienne utilisant des poids de pertinence clinique dans le cas de petits échantillons - Application au cancer du colon	33
5	Intégration de l'élicitation d'experts dans une méthode de sélection de variables en Bayésien par la méthode du <i>power prior</i> - Application au cancer du colon	55
	Synthèse	85
	Conclusion	95
	Perspectives	97
A	Compte-rendu de CHIMIO-ONCODIG de DxCare	101
B	Questionnaire portant sur l'élicitation des poids de sévérité	105
	Résumé court	121
	Abstract	123

## Liste des Figures

3.1	Exemple de l'ajustement de dose pour un patient en fonction de ses caractéristiques et des toxicités. . . . .	27
3.2	Exemple de l'évolution de la dose sur plusieurs cycles pour plusieurs patients. . . . .	28
3.3	Distributions de la dose et de la durée des cycles. . . . .	28
3.4	Distributions de l'âge, de la perte de poids, de la bilirubine, du score OMS et de la ligne de traitement sur l'ensemble des cycles. . . . .	31
3.5	Distribution des toxicités sur l'ensemble des cycles. . . . .	32
5.1	Densité de la distribution beta de $p$ de paramètres $a$ et $b$ et de moyenne $w/100$ . . . . .	89
5.2	Boxplot de la racine de l'erreur quadratique moyenne pour les méthodes LASSO, SSVS et WBS pour un scénario avec $n = 10$ patients et $K = 3$ cycles. . . . .	90
5.3	Pourcentage de sélection de chaque variable par chacune des méthodes LASSO, SSVS et WBS pour un scénario avec $n = 10$ patients et $K = 3$ cycles. Les variables utilisées pour simuler les données sont identifiées par des lignes en pointillé. . . . .	91
5.4	Flowchart de l'utilisation des patients pour le modèle SSVS couplé au <i>power prior</i> . . . . .	93
5.5	Pourcentage de sélection de chaque variable par le modèle SSVS couplé au <i>power prior</i> pour l'approche 1 et différentes valeurs de $a_0$ ( $a_0 \in \{0, 1; 0, 5; 1\}$ ). . . . .	93

## Liste des Tableaux

1.1	Exemple de classification de la CTCAE pour nos toxicités. . . . .	9
2.1	Pénalisations utilisées dans les principales méthodes de sélection de variables fréquentistes. . . . .	13
3.1	Effets indésirables rapportés avec l'irinotécan en traitement combiné (protocole de 180 mg/m <sup>2</sup> toutes les 2 semaines). . . . .	26
3.2	Statistiques descriptives sur l'ensemble des cycles. Les pourcentages sont donnés entre parenthèses. . . . .	29
3.3	Poids de pertinence clinique élicités par chaque clinicien pour chacune des variables. . . . .	30
5.1	Valeurs de $(a, b)$ intervenant dans la loi beta centrée sur les poids élicités/100. . . . .	88





# Production scientifique

## Articles publiés :

- Boulet S., Ursino M., Thall P., Jannot A.-S., Zohar S., Bayesian variable selection based on clinical relevance weights in small sample studies - Application to colon cancer. *Statistics in Medicine*. 2019 ;1-20. <https://doi.org/10.1002/sim.8107>
- Boulet S., Ursino M., Thall P., Landi B., Lepère C., Pernot S., Burgun A., Taieb J., Zaanani A., Zohar S., Jannot A.-S., Integration of elicited expert information via a power prior in Bayesian variable selection : application to colon cancer data. *Statistical Methods in Medical Research*. <https://doi.org/10.1177/0962280219841082>

## Conférences et colloques :

- 2017 :
  - Journées des moyens de calcul INSERM, Paris (France).  
Communication orale "Use case of UPMC cluster - Bayesian variable selection based on clinical relevance weights in small sample studies".
- 2018 :
  - ISCB ASC - Joint International Society for Clinical Biostatistics and Australian Statistical Conference, Melbourne (Australie).  
Communication orale "Bayesian variable selection based on clinical relevance weights in small sample studies - Application to colon cancer".
  - EPICLIN - Conférence Francophone d'Épidémiologie Clinique, Nice (France).  
Communication orale "Nouvelle méthode bayésienne de sélection de variables pour des échantillons de petite taille incorporant l'expertise clinique - Application au cancer colorectal".
  - ISBA - International Society for Bayesian Analysis, Edimbourg (Écosse).  
Poster "Bayesian variable selection based on clinical relevance weights in small sample studies - Application to colon cancer".
- 2019 :
  - ISBS - International Symposium on Biopharmaceutical Statistics, Kyoto (Japon).  
Communication orale "Integration of elicited expert information via a power prior in Bayesian variable selection : application to colon cancer data".
  - EPICLIN - Conférence Francophone d'Épidémiologie Clinique, Toulouse (France).  
Communication orale "Intégration de l'éllicitation d'experts dans une méthode de sélection de variables en Bayésien par la méthode de power prior - Application au cancer du colon".
  - DSVR - Colloque Données de santé en vie réelle, Paris (France).  
Communication orale " Intégration de l'éllicitation d'experts dans une méthode de sélection de variables en Bayésien par la méthode de power prior - Application au cancer du colon".
  - ISCB - International Society for Clinical Biostatistics, Louvain (Belgique).  
Poster "Integration of elicited expert information via a power prior in Bayesian variable selection".

**Collaboration internationale :**

- P. Thall, Department of Biostatistics, MD Anderson Cancer Center, Houston, Texas, USA.

# Glossaire

- AIC = *Akaike Information Criterion*.
- ARD = *Automatic Relevance Determination*.
- ATER = Attaché Temporaire d'Enseignement et de Recherche.
- BIC = *Bayesian Information Criterion*.
- CNRS = Centre National de la Recherche Scientifique.
- CRM = *Continual Reassessment Method*.
- CRC = Centre de Recherche des Cordeliers.
- CTCAE = *Common Terminology Criteria for Adverse Events* = Liste des effets indésirables classifiés par organes/symptômes qui peuvent être liés au traitement anti-cancer et qui permet de catégoriser les effets indésirables par des grades.
- DTR = *Dynamic Treatment Regimen*.
- GVS = *Gibbs Variable Selection* = Sélection de variables de Gibbs.
- HEGP = Hôpital Européen Georges Pompidou (Paris).
- Hb = Hémoglobine.
- iid = indépendantes, identiquement distribuées.
- INSERM = Institut Nationale de la Santé Et de la Recherche Médicale.
- GLM = *Generalized Linear Model* = Modèle Linéaire Généralisé.
- LARS = *Least Angle Regression*.
- LASSO = *Least Absolute Shrinkage and Selection Operator*.
- LMV = Laboratoire de Mathématiques de Versailles.
- MCMC = *Markov chain Monte-Carlo* = Méthode de Monte-Carlo par chaînes de Markov.
- MCU-PH = Maître de Conférences des Universités - Praticien Hospitalier.
- OMS = Organisation Mondiale de la Santé.
- OSCAR = *Octagonal Shrinkage and Clustering Algorithm for Regression*.
- PFS = *Progression Free Survival* = Survie sans progression.
- PP = *Power Prior*.
- RECORD = *REporting of studies Conducted using Observational Routinely-collected health Data*.
- SCAD = *Smoothly Clipped Absolute Deviation*.
- SSVS = *Stochastic Search Variable Selection*.
- TD = Travaux Dirigés.
- TTB = *Total Toxicity Burden*.
- UFR = Unité de Formation et de Recherche.
- UMR = Unité Mixte de Recherche.
- WBS = *Weight-Based SSVS*.

## Environnement de la thèse

Cette thèse a été réalisée au sein de l'équipe 22 "Sciences de l'information au service de la médecine personnalisée", UMRS 1138, de l'INSERM (Institut National de la Santé Et de la Recherche Médicale), dirigée par Anita Burgun, au Centre de Recherche des Cordeliers (CRC) à Paris, et au département Statistiques, Informatique et Santé Publique de l'Hôpital Européen Georges Pompidou (HEGP) d'où sont extraites les données utilisées dans cette thèse. J'ai été co-encadrée par Sarah Zohar, directrice de recherche en biostatistiques à l'INSERM, et Anne-Sophie Jannot, MCU-PH à l'HEGP.

Cette thèse a d'abord été financée trois ans par un contrat doctoral de l'Université Paris Descartes, puis un an par un contrat ATER (Attaché Temporaire d'Enseignement et de Recherche) à l'UFR de Sciences de l'Université Versailles/Saint-Quentin pour lequel j'ai rejoint l'équipe "Probabilités et Statistiques" dirigée par Catherine Donati-Martin, du Laboratoire de Mathématiques de Versailles (LMV), UMR 8100, du CNRS.

Durant mon contrat doctoral, j'ai mené pendant deux ans une mission d'enseignement au sein de l'UFR de Sciences de la Vie de l'Université Paris Diderot : TD de "Probabilités et biostatistiques" (L2S3), TD de "Statistiques appliquées à la biologie" (L2S4) et cours/TD d'"Initiation aux biostatistiques" (doctorants). J'ai ensuite pris en charge des cours/TDs de "Mathématiques générales" (L1S1), des TDs de "Probabilités & statistiques pour la biologie" (L1S2) et des TDs de "Mathématiques générales 2" (L1S2).

## Objectifs et déroulement de la thèse

Ma thèse porte sur l'intégration de données expertes dans des méthodes de sélection de variables dans le cadre Bayésien. Notre application consiste en la modélisation de la décision médicale de modification des doses de chimiothérapie en fonction des caractéristiques des patients et des effets indésirables.

Dans le cadre de notre premier projet, nous nous sommes placés dans le contexte des petits échantillons. La première méthode développée est une adaptation d'une méthode de sélection de variables Bayésienne dans laquelle des poids experts élicités sont introduits dans les distributions a priori.

Dans notre second projet, nous travaillons sur des échantillons plus importants et l'expertise clinique est introduite différemment dans le modèle. La deuxième méthode proposée est ainsi fondée sur la méthode de *power prior* et permet de combiner des données simulées à partir des poids élicités et les observations, dans l'objectif de sélectionner des variables importantes, mais rarement observées, et d'écartier des variables non pertinentes du point de vue expert, sans réduire les performances prédictives. Pour notre application, nous avons demandé à des oncologues experts dans le cancer colorectal d'éliciter des poids de pertinence clinique qui reflètent l'importance qu'ils donnent à chaque variable dans leur décision de réduire la dose, par l'intermédiaire d'un questionnaire qu'ils nous ont aidé à construire et qu'ils ont complété individuellement.

## Organisation du manuscrit

Dans un premier chapitre, nous introduisons le contexte statistique et clinique, et le rationnel sur la nécessité de sélectionner des variables en se basant à la fois sur les données cliniques et l'expérience d'experts. Dans le second chapitre, nous présentons les méthodes de sélection de variables existantes, à la fois fondées sur l'inférence fréquentiste et sur l'inférence Bayésienne, et la littérature disponible concernant l'élicitation de poids de sévérité selon l'expertise de cliniciens et leur utilisation. Puis, au chapitre 3, nous détaillons notre cas d'usage des patients traités pour un cancer colorectal métastatique par une combinaison de molécules incluant l'irinotécan. Les chapitres 4 et 5 sont dédiés aux méthodes bayésiennes de sélection de variables que nous avons développées. Enfin, ce manuscrit s'achève sur la synthèse, la conclusion et les perspectives.

## Contexte et enjeux

### 1.1 Quels sont les enjeux actuels des méthodes de sélection de variables pour les données de soin ?

Les innovations récentes concernant l’accessibilité et la disponibilité des dossiers de santé informatisés hospitaliers, rendent cette source de données structurées utilisable pour les analyses statistiques et la recherche médicale [35, 99]. La réutilisation des dossiers de santé informatisés peut permettre une meilleure compréhension de la prise de décision associée aux actions de traitement de type séquentiel. Néanmoins, la modélisation des décisions médicales derrière le choix des traitements comme une fonction des caractéristiques des patients, des traitements et doses précédents, et des résultats cliniques, implique de savoir prendre en compte un grand nombre potentiel de variables explicatives.

En effet, en pratique clinique, à chaque visite du patient, le médecin peut choisir un nouveau traitement, arrêter temporairement le traitement en cours ou modifier la dose en se basant sur son expérience, des recommandations externes, et les données médicales actuelles du patient. Le médecin dispose de beaucoup de variables mais il ne prendra sa décision qu’à partir d’un tout petit nombre d’entre elles en fonction de la maladie.

Les modèles de sélection de variables sont intéressants car ils peuvent permettre de sélectionner des variables pertinentes pour la prise de décision de façon automatisée. Sélectionner de telles variables pourraient permettre, dans une seconde étape, de construire des outils d’aide à la décision à destination des médecins pour optimiser les ajustements de doses afin d’obtenir le meilleur bénéfice en termes de toxicités/efficacité. L’ensemble des variables sélectionnées doit donc être réellement utilisable ou refléter la pratique clinique. Par conséquent, le nombre de variables doit être limité afin que (1) d’abord, le médecin ait le temps nécessaire pour récupérer toute l’information demandée et la rentrer dans le logiciel d’aide à la décision et que (2) le modèle puisse être compréhensible et interprétable par les cliniciens. Le second point implique que nous ne cherchons pas à développer de modèles extrêmement complexes de type boîtes noires, comme les réseaux de neurones, où le médecin ne sait pas quelles variables sont réellement utilisées pour réduire la dose et de quelle manière, de sorte qu’il préférera attribuer une dose à son patient à partir de son unique expérience plutôt que de prendre en compte la recommandation de l’outil d’aide à la décision proposé.

Par ailleurs, un algorithme d’apprentissage automatique basé uniquement sur les données des dossiers de santé informatisés hospitaliers risque d’omettre des variables rares mais importantes en terme d’actions thérapeutiques en raison d’un manque de puissance. De plus, une limite dans l’utilisation des données cliniques est qu’elles peuvent refléter des biais cognitifs comme des discriminations [16]. Par exemple, Deborde et al. ont montré que lorsque les médecins choisissent un traitement pour soigner l’hypertension, ils ont tendance à ne pas choisir le même traitement pour soigner les hommes et les femmes à comorbidités équivalentes, c’est-à-dire pour des hommes et des femmes qui se ressemblent

en tout point sur les autres critères, alors même qu’aucune recommandation n’indique que les hommes et les femmes doivent être soignés différemment [27].

Une source d’information complémentaire aux données est l’expertise clinique sur les variables qui doivent intervenir dans la décision d’après les experts. Cette expertise se trouve dans les recommandations professionnelles. Cependant, ces recommandations de traitement, sur lesquelles peuvent se baser les médecins, sont souvent issues des résultats d’essais cliniques, dans lesquels les patients constituent un sous-ensemble particulier de la population traitée réelle. Les critères d’inclusion et d’exclusion des essais cliniques peuvent être tellement stricts que des patients trop âgés, trop jeunes ou présentant des comorbidités particulières et multiples, seront exclus.

## 1.2 Comment modéliser la décision médicale à partir d’informations expertes ?

Un exemple de modélisation de l’information experte se trouve dans les systèmes d’aide à la décision clinique. Ces systèmes fournissent des algorithmes explicites incorporant des capacités de raisonnement basées sur des moteurs d’inférence et des bases de connaissances, y compris des règles validées, des directives partagées et des ontologies. Ils peuvent permettre de personnaliser la décision médicale en fonction des caractéristiques du patient.

Par exemple, Antibioclic est un outil d’aide à la décision, accessible en ligne, très utilisé par les médecins (<https://antibioclic.com>). Selon le domaine anatomique, la pathologie et les caractéristiques du patient, renseignés par le médecin, Antibioclic va fournir la molécule d’antibiothérapie qu’il faudrait utiliser d’après les recommandations des experts.

Cependant, d’après Smellie et al., les directives fournies par ces outils sont principalement axées sur une pathologie particulière et fournissent des conseils (parfois très) limités pour une interprétation spécifique au patient [86]. Le problème majeur des modèles, sur lesquels reposent ces outils d’aide à la décision, est qu’ils ont été construits à partir d’un petit nombre d’experts. Ils incorporent donc peu d’informations et peuvent être biaisés si les médecins ont construit leur algorithme à partir de patients qui ne sont pas représentatifs de la population de patients traités en soin courant. De plus, une variabilité des pratiques peut être observée, en partie à cause du pays d’origine, de la culture mais aussi de l’expérience personnelle du médecin et de la relation unique qui le lie à son patient.

En pratique clinique, le médecin ne va donc pas toujours suivre la recommandation proposée. Il va plutôt adapter au mieux sa décision à la situation clinique qui se présente à lui, en se basant sur un plus grand nombre de variables incluant les nombreuses caractéristiques du patient et sa propre expérience. La modélisation des décisions médicales dans les outils d’aide à la décision à base de connaissances est éloignée de la vie réelle.

Pour prendre en compte l’expertise clinique, une autre solution consiste à demander à des médecins experts d’élucider des poids de pertinence clinique qui reflètent leur expérience médicale [101]. Quand un médecin prend une décision concernant le traitement d’un patient ou la modification de dose, il attribue implicitement plus de poids à certaines caractéristiques du patient et à certains résultats cliniques antérieurs qu’à d’autres. Cette pratique implique que la décision est influencée par l’expérience personnelle du médecin et ses connaissances concernant la pertinence clinique de certaines variables. Pour analyser nos données, nous avons demandé à plusieurs médecins experts d’élucider, un par un, des poids pour chaque variable des patients qui est liée à leur décision médicale pour une indication et un traitement spécifique, incluant les toxicités.

## 1.3 Intérêt de la modélisation des décisions médicales en chimiothérapie

L'un des domaines où les patients en vie réelle sont particulièrement éloignés des patients étudiés dans les essais cliniques est le domaine de la chimiothérapie. Ainsi, les oncologues doivent adapter les recommandations issues de ces essais pour leur prise de décision. Par conséquent, comprendre et modéliser la décision médicale en chimiothérapie est une étape essentielle pour pouvoir dans le futur proposer des outils d'aide à la décision qui permettraient d'individualiser les traitements et d'optimiser le bénéfice en termes d'efficacité/toxicité pour un patient donné.

Nous nous plaçons dans un contexte où des doses sont ajustées sur plusieurs cycles pour des patients traités pour un cancer colorectal. En oncologie, un cycle est une série de traitement répété de manière régulière, avec des périodes de pauses entre chaque. Par conséquent, un nouveau cycle débute quand le patient reçoit une nouvelle dose. Dans le contexte du cancer colorectal, certains biomarqueurs pouvant expliquer des toxicités ont récemment été identifiés. Par exemple, il a récemment été montré que des mutations actives de l'oncogène KRAS étaient prédictives de la résistance des agents anti-EGFR, tels que le Cetuximab, qui est parfois utilisé en même temps que l'irinotécan dans le traitement du cancer colorectal métastatique [66]. Un autre exemple est le gène UGT1A1, qui est fortement associé au niveau de bilirubine, et pour lequel une mutation est liée à un risque accru de toxicité pour l'irinotécan. Ces résultats suggèrent que les sous-groupes de patients définis par ces variables devraient être traités différemment.

Les effets indésirables les plus communément associés à l'irinotécan incluent les vomissements, les nausées, les diarrhées, l'asthénie, la neutropénie et l'anémie. Toutes ces toxicités peuvent apparaître à des niveaux de sévérité différents, classifiés en grades 0, 1, 2, 3, 4 ou 5 selon la *Common Terminology Criteria for Adverse Events* (CTCAE) :

- Grade 0 : aucun effet indésirable ou normal ;
- Grade 1 : Léger ; asymptomatique ou symptômes légers ; diagnostic à l'examen clinique uniquement ; ne nécessitant pas de traitement ;
- Grade 2 : Modéré ; nécessitant un traitement minimal, local ou non-invasif ; interférant avec les activités instrumentales de la vie quotidienne ;
- Grade 3 : Sévère ou médicalement significatif mais sans mise en jeu immédiate du pronostic vital ; indication d'hospitalisation ou de prolongation d'hospitalisation ; invalidant ; interférant avec les activités élémentaires de la vie quotidienne ;
- Grade 4 : Mise en jeu du pronostic vital ; nécessitant une prise en charge en urgence ;
- Grade 5 : Décès lié à l'effet indésirable.

À titre d'exemple, le tableau 1.1 montre la classification de la CTCAE associée à quelques unes de ces toxicités.

Généralement, la dose initiale de chaque patient est calculée selon son poids, sa taille, son âge et ses marqueurs génétiques. Par la suite, des variables telles que la perte de poids, l'état de santé général et le niveau d'activité du patient (résumés par le score OMS), la ligne de traitement, la réponse au traitement, les polymorphismes pharmacogénétiques et les toxicités peuvent influencer la décision médicale de réduire la dose, d'allonger la durée entre deux intervalles d'administration de doses (appelés cycles) ou de choisir une autre ligne de traitement (protocole). L'analyse post-hoc d'essais cliniques randomisés et de données du monde réel a déjà mis en évidence une fréquence élevée d'ajustement de la posologie en fonction de l'âge avancé du patient, de comorbidités telles que le diabète, l'obésité [109, 11, 91]. De plus, l'âge, le score OMS, la bilirubine, le polymorphisme génétique et le schéma d'administration de la molécule peuvent être liés aux toxicités relatives à l'irinotécan [57]. Alors que des protocoles d'ajustement de doses de chimiothérapie en fonction des effets indésirables existent, ces recommandations ne tiennent pas compte des associations possibles entre les variables précédemment citées. C'est



pourquoi, en pratique clinique, les oncologues adaptent les doses et les calendriers d'attribution des doses au cours des cycles de traitement successifs en prenant en compte toutes les caractéristiques des patients et ses antécédents cliniques, incluant les doses et les résultats au traitement. Par conséquent, chaque régime de doses de chaque patient sur plusieurs cycles diffère de ce qui est recommandé dans les protocoles. Cela implique que les décisions sont des décisions personnalisées et que seules quelques observations parmi de nombreuses variables vont être disponibles dans les données cliniques. Par conséquent, en utilisant des stratégies basées sur les données, les variables rares ayant un fort impact sur les décisions médicales peuvent ne pas être sélectionnées.

## 1.4 Problématique de la thèse

Plutôt que d'utiliser uniquement les données cliniques ou les connaissances des experts pour modéliser la décision médicale, une alternative intéressante serait d'utiliser ces deux types d'information dans une même modélisation. L'objectif de cette thèse est donc de proposer de nouvelles méthodes de sélection de variables Bayésiennes permettant de combiner automatiquement données de vie réelle et données expertes.

Pour cela, nous avons développé deux méthodes qui combinent cette expertise avec les données observées pour comprendre le processus décisionnel entre effets indésirables et réductions de doses et nous avons appliqué ces méthodes à des données de patients traités par irinotécan pour un cancer colorectal. Le premier modèle se place dans le contexte des petits échantillons. En oncologie, beaucoup de nouvelles molécules et de nouveaux biomarqueurs (génomiques, cliniques ou physiologiques) ont récemment été proposés ce qui nous amène à devoir considérer beaucoup de petites sous-populations de maladies pouvant être supposées homogènes [3]. Dans de telles petites sous-populations, l'identification des causes des toxicités qui résultent en des modifications du traitement dans la prise en charge des patients peut être très difficile. Dans le contexte de petits échantillons, l'ajustement de modèles multivariés peut entraîner l'apparition de variances importantes pour les paramètres estimés, limitant la fiabilité inférentielle. Les méthodes de sélection de variables dans de telles situations peuvent atténuer cette difficulté en conservant uniquement les variables ayant les plus forts effets. Le modèle que nous proposons est un modèle Bayésien, construit à partir du modèle de *Stochastic Search Variable Selection* (SSVS), que nous avons appelé *Weight-Based SSVS* (WBS), où les paramètres des distributions a priori dépendent des poids de pertinence clinique élicités par les cliniciens experts [38]. Nous avons montré que la méthode développée avait de meilleures performances que les méthodes LASSO (*Least Absolute Shrinkage and Selection Operator*) et SSVS usuelles.

Pour le second modèle, des données simulées sont générées à partir des poids élicités par les experts, puis associées aux réductions de doses observées grâce à la méthode de *power prior* [50]. Nous avons montré que notre méthode permet en particulier de sélectionner des variables rares, qui ne seraient pas sélectionnées si nous n'utilisions que les données observées, et de mettre de côté des variables qui paraissent pertinentes à partir des données observées seules mais non du point de vue expert.

Tableau 1.1 – Exemple de classification de la CTCAE pour nos toxicités.

Toxicité	Type de toxicité	Définition	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5
Vomissements	Affections gastro-intestinales	Trouble caractérisé par l'acte réflexe de rejeter le contenu de l'estomac par la bouche.	1-2 épisodes (espacés de 5 minutes) en 24h	3-5 épisodes (espacés de 5 minutes) en 24h	>= 6 épisodes (espacés de 5 minutes) en 24h ; nécessitant une sonde de nutrition entérale, une nutrition parentérale totale ou une hospitalisation	Mise en jeu du pronostic vital ; nécessitant une prise en charge en urgence	Décès
Nausées	Affections gastro-intestinales	Trouble caractérisé par une sensation d'écœurement et/ou l'envie impérieuse de vomir	Perte de l'appétit sans modification des habitudes alimentaires.	Diminution des prises alimentaires sans perte significative de poids, déshydratation ou malnutrition	Apport calorique oral et apport hydrique insuffisants ; nécessitant une sonde de nutrition entérale, une nutrition parentérale totale ou une hospitalisation	-	-
Diarrhées	Affections gastro-intestinales	Trouble caractérisé par la survenue fréquente de selles liquides.	Augmentation de 4 ou moins du nombre de selles par jour par rapport à l'état initial ; légère augmentation des volumes de stomie par rapport à l'état initial	Augmentation de 4 à 6 du nombre de selles par jour par rapport à l'état initial ; augmentation modérée des volumes de stomie par rapport à l'état initial	Augmentation de 7 ou plus du nombre de selles par jour par rapport à l'état initial ; incontinence ; hospitalisation requise ; augmentation sévère des volumes de stomie par rapport à l'état initial ; interférant avec les activités élémentaires de la vie quotidienne	Mise en jeu du pronostic vital ; nécessitant une prise en charge en urgence	Décès
Anémie	Affections hématologiques et du système lymphatique	Trouble caractérisé par une diminution de la quantité d'hémoglobine dans 100 mL de sang. Les signes et symptômes d'une anémie peuvent comporter une pâleur de la peau et des muqueuses, un essoufflement, des palpitations cardiaques, un murmure systolique, une léthargie et une fatigabilité	Hémoglobine (Hb) < LIN - 10,0 g/dL ; < LIN - 6,2 mmol/L ; < LIN - 100 g/L	Hb < 10,0 - 8,0 g/dL ; < 6,2 - 4,9 mmol/L ; < 100 - 80g/L	Hb < 8,0 - 6,5 g/dL ; < 4,9 - 4,0 mmol/L ; < 80 - 65 g/L ; transfusion indiquée	Mise en jeu du pronostic vital ; nécessitant une prise en charge en urgence	Décès



## État des connaissances

### 2.1 Méthodes de sélection de variables

**Notations** Les notations suivantes sont introduites dans le cadre général de la sélection de variable. Soit  $n$  le nombre de sujets,  $J$  le nombre de variables, et  $K_i$  le nombre de cycles de chaque patient  $i$ ,  $i \in \{1, \dots, n\}$ . Pour un patient  $i \in \{1, \dots, n\}$ , notons  $\mathbf{x}_{i,k}$  le vecteur de taille  $J$  des variables disponibles au début du cycle  $k \in \{1, \dots, K_i\}$ . Dans notre application, certaines variables, comme l'âge et le traitement, sont supposées fixées dans le temps, tandis que d'autres caractéristiques des patients, des toxicités liées aux effets indésirables et les réponses aux traitements peuvent évoluer au cours du temps. Pour  $j \in \{1, \dots, J\}$ ,  $x_{i,k,j}$  prend ses valeurs dans  $\{x'_{j,0}, x'_{j,1}, \dots, x'_{j,C_j}\}$ , où  $x'_{j,c}$  est le cème niveau de sévérité le plus élevé,  $c = 0, \dots, C_j$ . Par exemple, l'âge est dichotomisé et prend ses valeurs dans  $\{0, 1\}$ , où 0 signifie "age < 80", et 1 signifie "age  $\geq$  80". Chaque toxicité est catégorisée par des grades entiers entre 0 (pas d'effet indésirable) et 3 ou 4 (le plus sévère).

**Modèle linéaire généralisé à effets mixtes** Soit  $\mathbf{z}_{i,k}$  le vecteur correspondant aux variables catégorielles et/ou ordinales des patient inclus dans  $\mathbf{x}_{i,k}$ , codé à présent sous la forme binaire, c'est à dire que  $\mathbf{z}_{i,k}$  est de longueur  $L = \sum_{j=1}^J C_j$  avec chaque  $x_{i,k,j}$  découpé en  $C_j$  variables binaires, prenant chacune une valeur de 1 si la catégorie correspondante a été sélectionnée dans  $x_{i,k,j}$  et 0 sinon.

Le cadre général de la pratique clinique liée aux données d'irinotécan est l'adaptation répétée de la dose au cours des cycles successifs de traitement en fonction des derniers traitements, doses et résultats. Sur le premier cycle, une première dose est choisie à partir des caractéristiques initiales du patient, incluant l'âge, la bilirubine, le score OMS et éventuellement les traitements précédents. Puis, si des toxicités sont observées sur ce cycle, le médecin peut adapter la dose sur le cycle suivant. Dans le but d'expliquer les décisions des médecins à partir des variables mises à jour sur le dernier cycle, et étant donnée la nature longitudinale des données, nous utilisons un modèle linéaire généralisé (GLM) à effets mixtes. Soit  $Y_{i,k}$  la dose ou réduction de doses du patient  $i$  au cycle  $k$  et  $\mathbf{z}_{i,k}$  le vecteur des variables du patient sur ce cycle, pour  $i \in \{1, \dots, n\}$  et  $k \in \{1, \dots, K_i\}$ . Le modèle est donné par l'équation :

$$g(\mathbb{E}(Y_{i,k} | \mathbf{Z}_{i,k} = \mathbf{z}_{i,k})) = \eta(\mathbf{z}_{i,k}) = \theta_0 + \boldsymbol{\theta}^T \mathbf{z}_{i,k} + \gamma_i \quad (2.1)$$

où  $g$  est la fonction de lien entre la variable à expliquer  $\mathbf{Y}$  et les variables explicatives  $\mathbf{z}$ ;  $\boldsymbol{\theta}$  est un vecteur de paramètres de longueur  $L$ ;  $\gamma_1, \dots, \gamma_n$  sont les effets aléatoires indépendants et identiquement distribués (iid)  $\mathcal{N}(0, \sigma_\gamma^2)$ .

Par la suite, selon la nature de dose continue ou de réduction de dose binaire choisie pour la variable dépendante,  $Y_{i,k} | \mathbf{Z}_{i,k} = \mathbf{z}_{i,k}$  sera supposée suivre respectivement une loi normale ou une loi de Bernoulli et  $g$  sera choisi respectivement tel que  $g(\mu) = \mu$  (lien linéaire) ou  $g(\mu) = \text{logit}(\mu) = \log(\mu/(1 - \mu))$  (lien logit).

Avec cette formulation, la propriété de Markov est supposée conditionnellement aux effets aléatoires telle que la dose/réduction de dose au cycle  $k$  dépende seulement des variables issues du cycle précédent et collectées dans  $\mathbf{z}_{i,k}$ . Par conséquent,  $\mathbf{z}_{i,k}$  inclut les indicatrices des grades de chaque toxicité étant apparues au cycle  $k - 1$  et chaque valeur actualisée des variables initiales. Notons que les toxicités ne constituent pas les variables dépendantes de la régression, mais les variables explicatives.

Notre objectif est de déterminer un sous-ensemble,  $\mathcal{I}$ , du vecteur  $\mathbf{z}$  décrivant correctement les données. Pour l'indice de la variable  $l \in \{1, \dots, L\}$ , notons  $I_l$  l'indicatrice de sélection de la variable prenant la valeur

$$I_l = \begin{cases} 1 & \text{si } z_{.,l} \text{ est dans le modèle} \\ 0 & \text{sinon,} \end{cases}$$

où  $z_{.,l}$  fait référence à la  $l$ ème variable par commodité.

### 2.1.1 Inférence fréquentiste

La plupart des méthodes de sélection de variables sont fréquentistes. On distingue en particulier les procédures de sélection pas à pas et les régressions pénalisées.

#### Procédures de sélection pas à pas

Les méthodes de sélection pas à pas (ascendante, descendante et mixte) consistent à débiter avec le modèle comprenant soit toutes les variables explicatives, soit aucune, puis, à chaque étape, des variables sont ajoutées ou éliminées du modèle en fonction de la valeur d'un critère de sélection comme la  $p$ -valeur, le coefficient de détermination, l'AIC (*Akaike Information Criterion*) [2],

$$AIC = -2\mathcal{L} + 2L$$

avec  $\mathcal{L}$  le maximum de la log-vraisemblance et  $L$  le nombre de variables du modèle complet, ou le BIC (*Bayesian Information Criterion*) [83],

$$BIC = -2\mathcal{L} + L \ln(n)$$

avec  $n$  la taille de l'échantillon. Essentiellement, ces procédures de sélection pas à pas visent à trouver le meilleur compromis entre biais et variance afin de minimiser l'erreur quadratique moyenne pour la prédiction.

Ces méthodes, bien que relativement peu demandeuses en calcul, peuvent donner des résultats très variables car elles reposent sur un processus discret : une variable est soit conservée dans le modèle, soit supprimée. De petits changements dans les données peuvent mener à des modèles très différents [102]. La plupart des méthodes deviennent difficilement gérables lorsque  $L$  augmente.

#### Régressions régularisées

Pour effectuer une sélection de variables dans le cadre d'une régression, on peut utiliser des critères de pénalisation tels que les régressions *bridge*, LASSO, SCAD (*Smoothly Clipped Absolute Deviation*), *elastic net* ou OSCAR (*Octagonal Shrinkage and Clustering Algorithm for Regression*) [102, 33, 31, 118, 13]. Afin d'estimer les coefficients ( $\hat{\boldsymbol{\theta}}$ ), ces méthodes maximisent la log-vraisemblance marginale pénalisée :

$$P\mathcal{L}_\lambda(\boldsymbol{\theta}, \sigma_\epsilon^2, \sigma_\gamma) = \mathcal{L}(\boldsymbol{\theta}, \sigma_\epsilon^2, \sigma_\gamma) - p_\lambda(\boldsymbol{\theta})$$

où :

$$-\mathcal{L}(\boldsymbol{\theta}, \sigma_\epsilon^2, \sigma_\gamma) = -\frac{1}{2} \sum_{i=1}^n \{K_i \ln(2\pi) + \ln|X_i| + (\mathbf{y}_{i,\cdot} - \theta_0 - \boldsymbol{\theta}^T \mathbf{z}_{i,\cdot})^T X_i^{-1} (\mathbf{y}_{i,\cdot} - \theta_0 - \boldsymbol{\theta}^T \mathbf{z}_{i,\cdot})\} \text{ dans le}$$

- cas d'un modèle linéaire à effets mixtes, avec  $X_i = \mathbb{1}_{K_i} \sigma_\gamma^2 \mathbb{1}_{K_i}^T + \sigma_\epsilon^2 I_{K_i}$  ;
- $p_\lambda(\boldsymbol{\theta})$  est une pénalisation qui dépend de la méthode choisie (cf. Tableau 2.1) et  $\lambda$  un paramètre de régularisation positif déterminé par validation croisée.

Tableau 2.1 – Pénalisations utilisées dans les principales méthodes de sélection de variables fréquentistes.

Méthode	Pénalisation
<i>bridge</i>	$\lambda \sum_{l=1}^L  \theta_l ^\gamma$ avec $\lambda \geq 0$ et $\gamma \geq 1$
LASSO	$\lambda \sum_{l=1}^L  \theta_l $ avec $\lambda \geq 0$
SCAD	$\lambda \sum_{ \theta_l  \leq \lambda}  \theta_l  + \sum_{\lambda <  \theta_l  < \gamma\lambda} \frac{2\gamma\lambda \theta_l  - \theta_l^2 - \lambda^2}{2(\gamma-1)} + \sum_{ \theta_l  \geq \gamma\lambda} \frac{\lambda^2(\gamma+1)}{2}$ avec $\lambda \geq 0$ et $\gamma > 2$
<i>elastic net</i>	$\lambda((1-\alpha) \sum_{l=1}^L \theta_l^2 + \alpha \sum_{l=1}^L  \theta_l )$ avec $\lambda \geq 0$ et $0 \geq \alpha \geq 1$
OSCAR	$\lambda(\sum_{l=1}^L  \theta_l  + c \sum_{l < l'} \max\{ \theta_l ,  \theta_{l'} \})$ avec $c \geq 0$

On obtient alors :

$$(\hat{\boldsymbol{\theta}}, \hat{\sigma}_\epsilon^2, \hat{\sigma}_\gamma) = \underset{\boldsymbol{\theta}, \sigma_\epsilon^2, \sigma_\gamma}{\operatorname{argmax}} P\mathcal{L}_\lambda(\boldsymbol{\theta}, \sigma_\epsilon^2, \sigma_\gamma).$$

Les propriétés de chaque méthode dépendent de la pénalisation choisie.

**La régression *bridge*** La régression *bridge* admet deux cas particuliers particulièrement connus : pour  $\gamma = 1$ , on retrouve la régression LASSO et pour  $\gamma = 2$ , la régression *ridge*.

**Le LASSO** Le paramètre  $\lambda$  permet de contrôler le degré de parcimonie, c'est à dire le nombre de variables à inclure dans le modèle [102]. Ainsi, un  $\lambda$  égal à 0 correspond à l'estimateur des moindres carrés. Au contraire, de plus grandes valeurs de  $\lambda$  amènent à sélectionner moins de variables. La nature de la contrainte de type LASSO tend à estimer certains coefficients à 0 exactement, tout en sélectionnant les autres variables, et donne donc des modèles facilement interprétables. Néanmoins, le nombre de variables sélectionnées est borné par le nombre d'observations, et le LASSO ne parvient pas à faire de la sélection groupée en sélectionnant une variable d'un groupe et en estimant les autres à zéro.

De nouvelles méthodes ont été développées afin de remédier à de tels problèmes comme le *group* LASSO qui permet de sélectionner ou pas un groupe entier de variables [114].

**SCAD** La pénalité plus complexe du SCAD donne lieu d'une part à des modèles plus parcimonieux en pénalisant davantage les coefficients les plus faibles et, d'autre part, à des biais réduits pour les gros coefficients [33].

**L'*elastic net*** L'*elastic net* a été développé afin de s'affranchir de deux problématiques liées aux propriétés du LASSO : la restriction du nombre de variables pouvant être sélectionnées et l'absence de sélection groupée [118]. Un paramètre  $\alpha$  permet de pondérer les effets respectifs des pénalités LASSO (norme  $L_1$ ) et *ridge* (norme  $L_2$ ). Dans le cas extrême où  $\alpha = 0$  (respectivement  $\alpha = 1$ ), la méthode consiste en une régression *ridge* (respectivement LASSO). Le choix du paramètre  $\alpha$  entre

0 et 1 autorise donc la sélection de plus de  $n$  variables, tout en restant parcimonieuse. De plus, la pénalisation *ridge* permet le regroupement de variables corrélées. Néanmoins, les groupes ne sont pas clairement déterminés et une analyse supplémentaire doit être réalisée pour les définir.

**OSCAR** La méthode OSCAR permet une sélection groupée des variables [13]. Elle utilise comme pénalité la somme d'une pénalité de type LASSO et d'une pénalité faisant intervenir une norme  $L_\infty$  par paire entre les coefficients. Cette norme encourage les coefficients à être égaux. Les groupes sont par la suite définis comme les ensembles de variables dont les coefficients estimés sont identiques, et non comme les variables ayant une relation de corrélation. Le choix des valeurs de  $\lambda$  et  $c$  conditionnent le nombre de groupes créés. En effet, plus ces valeurs sont prises grandes, plus le nombre de groupe est restreint.

**Inconvénients des méthodes de régularisation** D'abord, la régression pénalisée constitue essentiellement un problème d'optimisation qui procure uniquement une estimation ponctuelle de  $\theta$ . Néanmoins, les statisticiens ont pour habitude d'utiliser des intervalles de confiance ou la p-valeur. Ce problème peut être résolu en appliquant un échantillonnage *bootstrap* mais celui-ci est particulièrement gourmand en calcul [102]. De plus, il a été montré que les estimations *bootstrap* de l'erreur type des estimations LASSO pouvaient être instables, et ne sont pas consistantes si les vrais coefficients sont nuls [63].

Ensuite, la régression pénalisée nécessite de choisir des paramètres. La méthode la plus utilisée pour cela est la validation croisée [25]. Cependant, cette méthode est particulièrement couteuse en calcul. Par ailleurs, elle est habituellement utilisée pour évaluer les performances prédictives du modèle, et non pour identifier des variables pertinentes.

Enfin, ces méthodes de sélection de variables ont été validées et sont associées à de bonnes propriétés de convergence pour de larges échantillons. Dans un contexte de petites échantillons et d'un nombre de variables important à considérer, ces méthodes ne parviennent pas à produire des résultats fiables.

### 2.1.2 Inférence Bayésienne

La régression Bayésienne est une approche alternative intéressante lorsque l'on travaille sur de petits échantillons. Soit  $\theta$  un paramètre fixé et  $X$  une variable aléatoire de densité  $f(\cdot|\theta)$ . L'analyse Bayésienne va combiner l'information a priori sur le paramètre  $\theta$  et l'information provenant de l'échantillon dans la distribution a posteriori de  $\theta$  sachant  $x$ . Ceci constitue l'élément clef de la statistique Bayésienne et fournit une alternative intéressante aux méthodes fréquentistes. La distribution conditionnelle de  $\theta$  sachant les observations de l'échantillon  $x$  est définie comme la distribution a posteriori de  $\theta$  sachant  $x$ , notée par  $\pi(\theta|x)$ . En appliquant le théorème de Bayes, on obtient :

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int_{\Theta} f(x|\theta)\pi(\theta)d\theta}.$$

La distribution a posteriori de  $\theta$  sachant  $x$ ,  $\pi(\theta|x)$  reflète l'information mise à jour sur  $\theta$  après avoir observé l'échantillon  $x$ . Le choix des distributions a priori est l'un des points cruciaux de l'analyse de données Bayésienne. Les opposants à cette approche critiquent souvent le caractère arbitraire du choix de ces distributions, tandis que les défenseurs de l'approche louent un moyen d'introduire de la flexibilité dans l'analyse. Utiliser une distribution a priori informative nécessite d'avoir à disposition de l'information historique ou subjective, mais aussi de savoir introduire cette information dans la distribution. Une telle distribution peut avoir un impact important sur la distribution a posteriori, en particulier pour de petits échantillons. Par dépit, des distributions a priori non informatives sont souvent utilisées. L'analyse bayésienne avec des distributions a priori non informatives préserve l'apparence de l'objectivité et est donc de plus en plus reconnue par les statisticiens adeptes de l'approche

fréquentiste.

Dans le contexte Bayésien, la sélection de variables peut être effectuée en utilisant des distributions a priori de type "*spike and slab*" qui permettent d'identifier et d'estimer les distributions a posteriori des paramètres de régression non nuls, tout en éliminant les variables qui ne semblent pas prédictives de la variable dépendante. O'Hara et Sillanpää proposent une revue de la littérature des méthodes de sélection de variables Bayésiennes [71].

### Modèles de sélection avec indicatrice

L'approche de sélection de variable la plus intuitive consiste à choisir le *slab*,  $\theta_l|I_l = 1$ , égal à  $\theta'_l$ , et le *spike*,  $\theta_l|I_l = 0$ , égal à 0. Cette approche a été dérivée en plusieurs modèles.

**Modèle de Kuo-Mallick** Le modèle de Kuo-Mallick suppose que l'indicatrice de sélection de variables  $I_l$  et  $\theta'_l$  sont indépendants a priori, c'est à dire que [61]

$$P(I_l, \theta'_l) = P(I_l)P(\theta'_l).$$

La distribution a priori de chaque coefficient de régression est un mélange d'une loi de Dirac en zéro (le "*spike*") et d'une composante absolument continue (le "*slab*"). La loi de Dirac en zéro représente les connaissances a priori que le coefficient correspondant soit égal à 0, c'est à dire que la variable ne soit pas pertinente pour prédire la variable dépendante. À chaque itération de l'algorithme MCMC, les variables prédictives sont soit exclues, soit incluses dans le modèle. Le modèle ne requiert que la spécification des distributions a priori usuelles pour les paramètres du modèle. Cependant, quand  $I_l = 0$ , le mélange sera mauvais si la distribution a priori de  $\theta'_l$  est trop vague.

**Sélection de variables de Gibbs (GVS)** Le second modèle appelée sélection de variables de Gibbs (GVS) a été introduit par Dellaportas et al. en 2002, suivant l'idée de Carlin et Chib (1995) [28, 17]. Les distributions a priori de l'indicatrice et des coefficients sont maintenant supposées dépendre l'une de l'autre, c'est-à-dire que

$$P(I_l, \theta'_l) = P(\theta'_l|I_l)P(I_l).$$

En effet,  $\theta'_l$  est supposé suivre une loi de mélange a priori :

$$\theta'_l|I_l \sim (1 - I_l)\mathcal{N}(\tilde{\mu}, S) + I_l\mathcal{N}(0, \tau^2),$$

où les constantes  $(\tilde{\mu}, S)$  doivent être définis par l'utilisateur, et  $\tau^2$  est la variance a priori de  $\theta'_l$ . L'avantage du second modèle par rapport au premier est qu'il permet davantage de mélange. Cependant, l'algorithme nécessite de calibrer  $(\tilde{\mu}, S)$ .

### Stochastic Search Variable Selection (SSVS)

Dans cette approche introduite par George et McCulloch pour des modèles de régression linéaires, le *spike* est une distribution concentrée autour de 0. Ici,  $\theta_l = \theta'_l$  et l'indicatrice joue sur la distribution a priori de  $\theta_l$ , i.e., [38]

$$P(I_l, \theta'_l) = P(\theta'_l|I_l)P(I_l).$$

La distribution a priori de chaque coefficient  $\theta'_l$  est supposée suivre une loi de mélange de deux distributions normales centrées en zéro mais avec des variances différentes, l'une grande et l'autre petite :

$$\theta'_l|I_l \sim (1 - I_l)\mathcal{N}(0, \tau^2) + I_l\mathcal{N}(0, g\tau^2) \tag{2.2}$$



Dans l'équation 2.2, le premier terme ("*spike*") est celui des variables non-sélectionnées ( $I = 0$ ) avec une densité centrée en 0 ayant une petite variance. On peut remarquer que la suite de distributions normales centrées en zéro converge vers la distribution de Dirac en 0 puisque sa variance tend vers 0. Cette approximation permet d'utiliser les logiciels de statistique usuels. Le second terme de l'équation, où  $I = 1$ , prend en compte les variables sélectionnées, et a donc une variance importante. Ce modèle requiert la spécification des paramètres a priori ( $\tau^2$  et  $g\tau^2$ ) qui dépendent des données (ou au moins du contexte). Contrairement au modèle GVS, les valeurs des paramètres quand  $I_l = 0$  influencent la distribution a posteriori.

Farcomeni et al. ont présenté une stratégie permettant de contraindre la sélection de variables. Ainsi, en plus des contraintes hiérarchiques et de regroupement usuelles, ils ont défini des contraintes anti-hiérarchiques où l'inclusion d'une variable en force une autre à être exclue du modèle [34].

Le modèle SSVS a été adapté à des modèles plus complexes comme les modèles pharmacocinétiques, la construction de portefeuilles d'actions en finance, des modèles linéaires généralisés, des modèles log-linéaires, des modèles probit multinomiaux et des modèles de régression multivariés [111, 39, 69, 84, 15]. Le modèle SSVS a été particulièrement appliqué dans le cadre génétique [92, 94, 112, 58, 70, 95, 59, 67, 97, 93, 96, 62, 1, 36, 90, 85, 29, 81].

Rockova a suggéré d'utiliser une loi de mélange de type *spike-and-slab* faisant intervenir une distribution de Laplace [80] :

$$\theta_l | I_l \sim (1 - I_l) \frac{\lambda_0}{2} \exp(-\lambda_0 |\theta_l|) + I_l \frac{\lambda_1}{2} \exp(-\lambda_1 |\theta_l|)$$

Ce modèle, connu sous le nom de *spike-and-slab* LASSO, utilise une suite de lois de mélange de type Laplace avec une pénalité *spike*  $\lambda_0$  qui augmente tandis que  $\lambda_1$  reste fixé à une valeur constante faible ( $\lambda_0 \gg \lambda_1$ ). Initialement développé dans le cadre d'une régression linéaire, il a été étendu aux modèles linéaires généralisés [98].

### Approximation de la forme "*spike and slab*"

Une autre approche consiste à ne pas utiliser d'indicatrices de sélection  $I_l$  dans le modèle, mais à spécifier une distribution a priori sur  $\theta_l$  qui donne une approximation de la forme "*spike and slab*". On utilise donc  $\theta_l = \theta'_l$ , avec comme distribution a priori

$$\theta'_l | \tau_l^2 \sim \mathcal{N}(0, \tau_l^2).$$

La distribution a priori de  $\tau_l^2$  doit être choisie de sorte à ce que les coefficients  $\theta'_l$  soient proches de 0 si la variable ne doit pas être sélectionnée. Inversement, les coefficients des variables sélectionnées ne doivent pas être sous-estimés. Le nombre de variables à sélectionner peut être ajusté en choisissant une distribution a priori de  $\tau_l^2$  appropriée. Cette approche n'utilise pas d'indicatrice de sélection de variable mais il reste possible d'en construire une en fixant par exemple un seuil  $c$ , tel que  $I_l = 1$  si  $|\theta'_l| > c$ .

**Distribution a priori Inverse Gamma** Une possibilité de distribution a priori pour  $\tau_l^2$  est une distribution Inverse Gamma :

$$\tau_l^2 \sim IG\left(\frac{a}{2}, \frac{2}{b}\right).$$

Les hyper-paramètres  $a$  et  $b$  sont habituellement choisis de manière à ce que la variance de  $\tau_l^2$  soit importante. Ce modèle est équivalent au modèle de *automatic relevance determination* (ARD) de Li et al. [64]. Un tel modèle est par exemple utilisé par Bae et Mallick qui étudie différents degrés de parcimonie à travers différentes distributions a priori et propose une application de sélection de variables à des données sur la leucémie et sur le cancer du sein [4].

**Distribution a priori de Jeffrey** Une autre distribution a priori pouvant être utilisée est la distribution a priori de Jeffrey  $\tau_l^2 \propto \frac{1}{\tau_l^2}$ , également utilisée dans [4].

**LASSO Bayésien** Le LASSO Bayésien propose d'utiliser une distribution a priori exponentielle pour  $\tau_l^2$  de paramètre  $\mu$  [73, 113]. La distribution a priori de  $\theta'_l|\mu$  est alors une distribution double exponentielle :

$$\theta'_l|\mu \propto \frac{\lambda}{2} \exp(-\lambda|\theta'_l|)$$

où  $\lambda$  est un paramètre qui dépend de la variance. Les coefficients non pertinents sont alors estimés à 0 exactement. Yi et Xu ont, par exemple, proposé une application aux données génétiques [113].

**Elastic Net Bayésien** Zou et Hastie ont montré en 2005 que la version Bayésienne de l'*elastic net* revient à trouver le mode marginal a posteriori de  $\theta'|\mathbf{y}$  quand la distribution a priori de  $\theta'$  est donnée par [118]

$$\theta' \propto \exp(-\lambda_1\|\theta'\|_1 - \lambda_2\|\theta'\|_2^2),$$

un compromis entre les distributions a priori Gaussiennes et de Laplace.

**Horseshoe prior** Carlos et al. ont introduit en 2010 le *horseshoe* [18] :

$$\pi_l|\lambda \sim C^+(0, \lambda), \lambda|\sigma \sim C^+(0, \sigma)$$

où  $C^+(0, \lambda)$  est une distribution *half*-Cauchy standard sur les réels positifs avec comme paramètre d'échelle  $\lambda$ .

### L'approche de l'espace modèle

Les modèles précédents ont pour point commun d'attribuer des distributions a priori aux coefficients  $\theta_l$ . Une autre approche consiste à attribuer des distributions a priori au nombre de variables sélectionnées  $N_v$  et aux coefficients associés  $(\theta_1, \dots, \theta_{N_v})$ . Le nombre de variables sélectionnées est une variable aléatoire, et peut être contrôlé à travers la distribution a priori de  $N_v$ . Parmi les modèles utilisant une telle approche, on peut citer le *Reversible jump MCMC* et le *Composite model space (CMS)* [42].

Dans les chapitres 4 et 5, les variables sont sélectionnées à l'aide de la méthode SSVS. Cette méthode Bayésienne de sélection de variables autorise, entre autres, à introduire de l'information experte dans ses distributions a priori.

## 2.2 Comment représenter l'information experte ? Principe de l'élicitation.

Pour introduire de l'information experte dans un modèle Bayésien de sélection de variables, il est nécessaire de définir une méthodologie permettant de résumer cette information. Dans la littérature, l'élicitation de poids de sévérité est surtout utilisée en méthodologie de recherche de doses [6, 23, 32]. Par exemple, en 2004, Bekele et Thall utilisent des poids de sévérité élicités dans un contexte de recherche de dose pour un essai de phase I d'une molécule de chimiothérapie [6]. Le but scientifique d'un essai de phase I en oncologie d'un nouvel agent chimiothérapeutique est de trouver une dose présentant un niveau de toxicité acceptable. Pour des raisons éthiques, la recherche de la dose est faite de manière adaptative, les doses étant choisies pour des cohortes successives de patients sur la base des données obtenues des cohortes précédentes. Généralement, les patients sont exposés à plusieurs

types de toxicité pouvant prendre différents grades de sévérité. Les oncologues qui planifiaient l'essai voulaient tenir compte des différences de sévérité entre les toxicités qu'ils avaient identifiées. Ils ont également demandé que la méthode de recherche de dose tienne compte du fait qu'une toxicité de faible grade observée à une dose donnée, bien que ne donnant pas lieu à une limitation de la dose, soit perçue comme un avertissement qu'un grade de toxicité supérieur est susceptible d'apparaître à une dose plus élevée. Les méthodes conventionnelles de phase I réduisent chaque type de toxicité à une indicatrice de la survenue d'un grade de sévérité considéré comme limitant la dose, définissent la variable "toxicité" comme le maximum de ces indicatrices et fondent la recherche de dose sur cette seule variable binaire. Comme les méthodes conventionnelles ne répondaient pas aux préoccupations des médecins, Bekele et Thall ont mis au point une méthode bayésienne de recherche de dose utilisant un vecteur de toxicités corrélées, à valeurs ordinales, avec des niveaux de sévérité variant avec la dose. Ils ont également mis au point une méthode permettant d'éliciter conjointement la distribution a priori, un vecteur de poids quantifiant l'importance clinique de chaque grade de chaque type de toxicité et une toxicité totale cible (*target Total Toxicity Burden -TTB*) acceptable pour les médecins. Ainsi, les médecins ont attribué un poids de sévérité pour chaque grade de chaque type de toxicité dans l'intervalle allant de 0 (aucune importance clinique) à 10 (très forte importance clinique). Le TTB a ensuite été défini comme la somme sur les différents grades et types de toxicités pondérés par les poids de sévérité élicités par les cliniciens. La méthode attribue à chaque cohorte la dose correspondant au TTB moyen a posteriori le plus proche de la cible. Le processus d'élicitation est itératif : Les résultats de l'algorithme sont régulièrement présentés aux oncologues qui sont autorisés à modifier les poids qu'ils ont précédemment choisis.

Ezzalfani et al. se sont inspiré de ce travail en réutilisant de tels poids de sévérité pour construire un nouveau score de toxicité, appelé TTP (*Total Toxicity Profile*), défini comme la norme euclidienne des poids [32]. Cette modification leur permet de différencier des situations cliniquement très différentes, qui étaient traitées de la même manière par le TTB. Le TTP est ensuite introduit dans la méthode CRM (*continual reassessment method*), avec un modèle logistique pour modéliser la relation dose-toxicité dans un cadre fréquentiste, afin de proposer une nouvelle méthode de recherche de doses. En 2006, Yuan et al. [115] ont également utilisé un score de toxicité qui prend en compte la sévérité des grades de toxicité dans la procédure d'allocation des doses de la CRM en utilisant une vraisemblance quasi-Bernoulli dans un cadre Bayésien.

Une approche différente de la recherche de doses en phases I-II utilise les utilités élicitées des paires possibles de résultats (efficacité, toxicité) [48, 100].

## 2.3 Comment introduire l'information experte dans les modèles utilisant l'inférence Bayésienne ?

L'élicitation de distribution a priori informative est l'un des sujets les plus importants de l'inférence bayésienne, étant de plus en plus utilisée à l'ère du *Big Data* dans des cadres tels que les essais cliniques et les études observationnelles. L'élicitation de l'information a priori n'est généralement pas une tâche facile car elle nécessite de quantifier et de synthétiser l'information a priori dans une distribution appropriée. Ainsi, le développement de nouvelles techniques est indispensable. Actuellement, il existe des méthodes pour introduire de l'information externe mais pas des données expertes élicitées.

### 2.3.1 Introduction dans les distributions a priori

Un avantage majeure de l'inférence Bayésienne est que de l'information externe peut être introduite formellement dans les distributions a priori. De nombreuses études bayésiennes portant sur la sélection de variables utilisent des distributions a priori qui ne spécifient pas de préférence pour une variable, mais peuvent, par exemple, indiquer une préférence pour des modèles parcimonieux. D'autres études

ont proposé des approches simples pour introduire les connaissances a priori, en attribuant indépendamment à chaque variable une probabilité a priori d'être incluse dans le modèle [24]. Par exemple, Kitchen et al. ont fourni une méthode de spécification des paramètres a priori utilisant la littérature scientifique pour les problèmes de régression avec de nombreux prédicteurs dans lesquels les distributions a priori incluent des composantes de réduction et de sélection de variables, étendant le modèle de Kuo et de Mallick [61]. Les auteurs proposent cette approche comme une solution au problème phénotype-génotype. Ils construisent une distribution a priori en attribuant un nombre à chaque position du codon en fonction du nombre de fois où il est identifié comme important dans la littérature [60].

À notre connaissance, les méthodes de sélection des variables introduisant des distributions a priori informatives ont principalement été considérées dans des contextes génomiques. Dans ce contexte, les distributions a priori sont établies en exploitant une vaste littérature sur la sévérité des variants génétiques, tant du point de vue expérimental que bioinformatique. Par conséquent, dans les contextes génétiques, la variabilité des distributions a priori est considérée comme faible et les études de simulation sur la variabilité des performances dues à une spécification antérieure ont suscité peu d'intérêt [60]. Ce n'est pas le cas des covariables extraites des dossiers de santé informatisés, pour lesquelles la littérature est extrêmement limitée en ce qui concerne leurs effets respectifs sur la gravité de la maladie et donc leur importance dans la prise de décision médicale.

Au chapitre 4, nous proposons d'utiliser l'information experte modélisée par des poids élicités pour construire les hyperparamètres de la distribution a priori de l'indicatrice de sélection de variables intervenant dans la méthode SSVS.

### 2.3.2 Power Prior

#### Définition

En présence de données historiques, l'élicitation de distributions a priori informatives paraît plus "objective" dans un certain sens. Le *Power Prior* (PP) défini par Ibrahim et Chen permet d'intégrer de l'information contenue dans les données historiques dans l'analyse et l'inférence Bayésiennes d'une étude actuelle [50]. Le PP est défini comme le produit de la vraisemblance pondérée de  $\theta$  conditionnellement aux données historiques et d'une distribution a priori non-informative :

$$\pi(\theta|\mathbf{D}_0, a_0) \propto L(\theta|\mathbf{D}_0)^{a_0} \pi_0(\theta) \quad (2.3)$$

où  $\mathbf{D}_0 = (n_0, \mathbf{y}_0, \mathbf{z}_0)$  correspond aux données historiques,  $L(\theta|\mathbf{D}_0)$  est la vraisemblance de  $\theta$  conditionnellement aux données historiques,  $0 \leq a_0 \leq 1$  est un paramètre d'échelle qui contrôle le poids attribué aux données historiques dans l'analyse, et  $\pi_0(\theta)$  est une distribution non-informative pour  $\theta$ . Si le paramètre du power prior  $a_0$  est égal à zéro, aucune information historique n'est utilisée et la distribution a priori est considérée comme non informative. La distribution a posteriori de  $\theta$  s'écrit :

$$\begin{aligned} \pi(\theta|\mathbf{D}, \mathbf{D}_0, a_0) &\propto L(\theta|\mathbf{D})\pi(\theta|\mathbf{D}_0, a_0) \\ &\propto L(\theta|\mathbf{D})L(\theta|\mathbf{D}_0)^{a_0}\pi_0(\theta) \end{aligned}$$

où  $\mathbf{D} = (n, \mathbf{y}, \mathbf{z})$  sont les données observées et  $L(\theta|\mathbf{D})$  est la vraisemblance de  $\theta$  conditionnellement aux données observées.

### Extension à un paramètre de PP aléatoire

Le PP de l'équation 2.3 a initialement été défini pour  $a_0$  fixé. Chen et al. ont étendu le modèle au cas où  $a_0$  est aléatoire [21] :

$$\pi(\boldsymbol{\theta}, a_0 | \mathbf{D}_0) \propto L(\boldsymbol{\theta} | \mathbf{D}_0)^{a_0} \pi_0(\boldsymbol{\theta}) \pi(a_0 | \gamma_0) \quad (2.4)$$

Un paramètre  $a_0$  aléatoire donne au statisticien plus de flexibilité dans le poids à accorder aux données historiques. Une distribution Beta ou une distribution uniforme constituent des distributions a priori naturelles pour  $a_0$ .

### Extension à plusieurs jeux de données historiques

Ibrahim et Chen ont également généralisé la distribution a priori donnée par l'équation 2.4 à plusieurs jeux de données historiques [50]. Supposons qu'il y ait  $M$  jeux de données historiques disponibles. Soit  $\mathbf{D}_{0,m}$  le  $m$ ème jeux de données historiques,  $m \in \{1, \dots, M\}$ , et  $\mathbf{D}_0 = (\mathbf{D}_{0,1}, \dots, \mathbf{D}_{0,M})$ . Pour  $m \in \{1, \dots, M\}$ , on définit  $a_{0,m}$  le paramètre de PP tel que les  $a_{0,m}$  soient des variables aléatoires indépendantes et identiquement distribuées selon une loi Beta d'hyperparamètres  $(\alpha, \beta)$ . Soit  $\mathbf{a}_0 = (a_{0,1}, \dots, a_{0,M})$ . Le PP pour plusieurs jeux de données historiques a été défini comme :

$$\pi(\boldsymbol{\theta}, \mathbf{a}_0 | \mathbf{D}_0) \propto \left[ \prod_{m=1}^M L(\boldsymbol{\theta} | \mathbf{D}_{0,m})^{a_{0,m}} \pi(a_{0,m} | \alpha, \beta) \right] \pi_0(\boldsymbol{\theta})$$

Ibrahim et Chen ont formalisé l'utilisation du PP comme distribution a priori pour de nombreux modèles de régression [50]. Le PP peut en effet être utilisé dans des cadres tels que ceux des modèles linéaires généralisés, des modèles mixtes linéaires généralisés, des modèles de survie avec données censurées, des modèles de fragilité, des modèles multivariés et des modèles non linéaires.

Le PP est objectif dans le sens où le degré d'information de la distribution a priori est déterminé par les informations contenues dans des données historiques. Le seul hyperparamètre nécessitant une élicitation subjective du PP est le paramètre  $a_0$ , pour lequel il est vivement recommandé de mener plusieurs analyses de sensibilité, notamment des analyses où  $a_0 = 0$  (a priori non informatif) et  $a_0 = 1$ . Une justification formelle du PP est donnée dans [53] où il est démontré qu'il s'agit d'une classe optimale d'a priori informatif dans le sens où elle minimise une somme convexe de divergences de Kullback-Leibler (KL) entre deux densités a posteriori spécifiques, l'une d'elle étant basée sur l'absence d'incorporation de données historiques, et l'autre sur la mise en commun des données historiques et actuelles. Ce résultat fournit une forte motivation pour utiliser le PP en tant qu'a priori informatif dans l'inférence bayésienne.

Le PP a déjà été utilisé dans les essais cliniques, en génétique, pour la santé mentale, en psychologie, dans la santé environnementale, en ingénierie, en économie et dans le commerce. De multiples exemples de l'utilisation du PP et de ses extensions dans les essais cliniques et les études sur le cancer peuvent être trouvés dans la littérature [8, 30, 44, 89, 9, 49, 54, 52, 22, 19, 50, 21, 53, 26, 10, 20, 46, 79, 51, 45, 107]. D'autres exemples utilisant le PP pour des études épidémiologiques incluent [43, 14, 37, 88].

Notre proposition du chapitre 5 consiste à remplacer les données historiques par des données expertes générées à partir de poids élicités.

## 2.4 Combinaison de plusieurs expertises

Les deux sections précédentes expliquaient comment résumer l'information experte et comment l'introduire dans des modèles Bayésiens. Néanmoins, la divergence des opinions d'experts peut nécessiter l'utilisation d'une méthodologie permettant de combiner les différentes expertises.

### 2.4.1 La méthode Delphi

Plusieurs méthodes, plus ou moins élaborées, peuvent permettre d'obtenir un consensus des opinions d'un ensemble d'expert. La technique la plus connue est la méthode Delphi [72]. Celle-ci permet d'éviter la confrontation directe des experts en les soumettant à des interrogatoires individuels répétés (en face à face ou par questionnaire) qui sont systématiquement suivis d'un compte-rendu contrôlé des opinions recueillies. Les questions sont choisies de manière à faire ressortir le raisonnement suivi par l'expert pour répondre à la question principale, les facteurs qu'il considère pertinents concernant le problème, sa propre estimation de ces facteurs et des informations sur le type de données qui, selon lui, permettraient d'arriver à une meilleure évaluation de ces facteurs, et de ce fait, à une réponse plus sûre à la question principale. L'information fournie aux experts entre deux séries de questionnement est généralement de deux types : Il s'agit soit des données disponibles précédemment demandées par l'un des experts, soit de facteurs et de considérations suggérés comme potentiellement pertinents par l'un ou l'autre des experts.

Ce mode d'interaction contrôlée entre les répondants permet d'éviter les inconvénients associés à des interactions directes. La méthode Delphi est plus propice à la conservation de l'opinion propre de l'expert et permet de l'aider dans la formation progressive d'un avis plus réfléchi. La confrontation directe, en revanche, induit trop souvent la formulation hâtive de notions préconçues, la tendance à se fermer aux nouvelles idées, celle de défendre la position prise et une prédisposition à se laisser influencer par les opinions des autres experts exprimées de manière plus convaincante.

En étudiant systématiquement les facteurs qui influencent le jugement de l'expert, il devient possible de corriger toute idée fautive qu'il aurait pu nourrir concernant des facteurs empiriques ou des hypothèses théoriques sous-jacentes à ces facteurs, et d'attirer son attention sur d'autres facteurs qu'il aurait peut-être négligés lors de sa première analyse de la situation. Les expérimentateurs doivent faire preuve d'une grande discrétion dans tout effort visant à faire changer d'avis un expert afin d'obtenir des résultats dénués de tout parti pris. Un moyen de s'assurer de cela est de n'introduire que les données demandées par au moins un expert et pouvant être obtenues à partir de sources fiables, et de suggérer uniquement les hypothèses théoriques qui semblent représenter le consensus de la majorité des répondants.

Si l'objectif de l'expérience est l'estimation d'une quantité numérique, on peut s'attendre à ce que, même si les avis initialement exprimés sont très divergents, les estimations individuelles montrent une tendance à convergence à mesure que l'expérience se poursuit.

### 2.4.2 Approche mathématique

Une solution alternative au consensus peut être l'utilisation d'approches mathématiques comme des moyennes, éventuellement pondérées, ou la méthode du *Bayesian Model Averaging* (BMA).

#### *Bayesian Model Averaging (BMA)*

La méthode BMA calcule la distribution prédictive d'un coefficient  $\theta_l$ ,  $l \in \{1, \dots, L\}$ , comme une moyenne pondérée des distributions a posteriori de  $\theta_l$  sous chaque modèle  $\mathcal{M}(m)$ ,  $m \in \{1, \dots, M\}$  [77] :

$$\mathbb{P}(\theta_l | \mathbf{D}) = \sum_{m=1}^M \mathbb{P}(\mathcal{M}(m) | \mathbf{D}) \mathbb{P}(\theta_l | \mathcal{M}(m), \mathbf{D})$$

où les poids  $\mathbb{P}(\mathcal{M}(m) | \mathbf{D})$  sont les probabilités a posteriori de chaque modèle :

$$\mathbb{P}(\mathcal{M}(m) | \mathbf{D}) = \frac{\mathbb{P}(\mathbf{D} | \mathcal{M}(m)) \mathbb{P}(\mathcal{M}(m))}{\sum_{m_1=1}^M \mathbb{P}(\mathbf{D} | \mathcal{M}(m_1)) \mathbb{P}(\mathcal{M}(m_1))}$$

La vraisemblance marginale du modèle  $\mathcal{M}(m)$  est :

$$\mathbb{P}(\mathbf{D}|\mathcal{M}(m)) = \int \mathbb{P}(\mathbf{D}|\boldsymbol{\theta}, \mathcal{M}(m))\mathbb{P}(\boldsymbol{\theta}|\mathcal{M}(m))d\boldsymbol{\theta}$$

où  $\mathbb{P}(\mathbf{D}|\boldsymbol{\theta}, \mathcal{M}(m))$  est la vraisemblance que nous pouvons évaluer en utilisant la méthode de Laplace et  $p(\boldsymbol{\theta}|\mathcal{M}(m))$  est la distribution a priori. L'espérance et la variance de  $\theta_l$  sont obtenus par :

$$\begin{cases} \mathbb{E}(\theta_l|\mathbf{D}) = \sum_{m=1}^4 \hat{\theta}_l^m \mathbb{P}(\mathcal{M}(m)|\mathbf{D}) \\ \mathbb{V}(\theta_l|\mathbf{D}) = \sum_{m=1}^4 [\mathbb{V}(\theta_l|\mathbf{D}, \mathcal{M}(m)) + (\hat{\theta}_l^m)^2] \mathbb{P}(\mathcal{M}(m)|\mathbf{D}) - \mathbb{E}(\theta_l|\mathbf{D})^2 \end{cases}$$

où  $\hat{\theta}_l^m = \mathbb{E}(\theta_l|\mathbf{D}, \mathcal{M}(m))$ .

La méthode BMA a été appliquée à de nombreux modèles statistiques incluant la régression linéaire, les modèles linéaires généralisés, les modèles de régression de Cox, les modèles graphiques et les études cas-témoins [78, 76, 110, 65, 108]. Cette méthode a aussi été utilisée dans des études d'association génétique et des études d'association des locus [5, 12].

### Loi de mélange

Pour combiner les opinions des experts qui peuvent diverger, une autre méthode peut consister en l'utilisation d'une loi de mélange obtenue en pondérant chacun des ensembles de poids cliniques élicités par chaque expert par un poids qui traduit la fiabilité que l'on accorde à cet expert :

$$\mathbb{P}(\theta_l|\mathbf{D}) = \sum_{m=1}^M \pi_m \mathbb{P}(\theta_l|\mathcal{M}(m), \mathbf{D})$$

avec  $\pi_m = \mathbb{P}(\mathcal{M}(m))$  tel que  $\pi_m \geq 0$  et  $\sum_{m=1}^M \pi_m = 1$ . Ces deux dernières contraintes assurent que cette combinaison linéaire de fonctions de densité est bien elle-même une fonction de densité.

Dans la littérature, les lois de mélange apparaissent plutôt lorsqu'une population statistique est composée d'au moins deux sous-populations. Dans ce cas, les poids  $\pi_m$ ,  $m \in \{1, \dots, M\}$ , correspondent à la proportion de chacune des sous-populations dans la population totale. Elles sont aussi parfois utilisées pour représenter des distributions non-normales.

Dans nos travaux, nous utilisons la méthode BMA et/ou la loi de mélange pour combiner les informations expertes obtenues de différents cliniciens.



# Présentation du cas d'usage

## 3.1 Le cancer colorectal

Les données suivantes sont issues du site de la ligue contre le cancer ([https://www.ligue-cancer.net/article/26096\\_cancer-du-colon-et-du-rectum#1](https://www.ligue-cancer.net/article/26096_cancer-du-colon-et-du-rectum#1)).

### 3.1.1 Généralités

En France, on compte près de 40 000 nouveaux cas de cancer colorectal chaque année, dont 53 % d'hommes et 47 % de femmes, en faisant le troisième cancer le plus fréquemment diagnostiqué, après ceux de la prostate et du sein. Il tue également 17 500 personnes par an, ce qui le place à la seconde place des cancers les plus mortel, derrière celui du poumon.

Des lésions bénignes, appelées polypes, peuvent en grossissant se développer en cancer. Détecter et supprimer les polypes les plus gros est donc essentiel pour diminuer le risque de développement d'un cancer colorectal. Enfin, le cancer colorectal touche dans 60 % des cas principalement le colon, et dans 40 % des cas le rectum.

### 3.1.2 Facteurs de risque

Bien que toutes les causes du cancer colorectal ne soient pas connues, les facteurs de risques suivants ont été identifiés.

#### Facteurs de risques internes

**L'âge** Le cancer colorectal touche en grande majorité les personnes de plus de 50 ans (95 % des cas, dont 46 % après 74 ans). Lorsqu'il se déclare de façon précoce, il est en général dû à un facteur génétique.

**La prédisposition génétique** Deux formes de cancers colorectaux impliquent un facteur génétique : la polyadénomatoase recto-colique familiale (mutation d'un gène nommé APC) et le syndrome de Lynch (anomalies sur des gènes codants pour les protéines de réparation de l'ADN). Ils constituent moins de 5 % de ces cancers et apparaissent avant 40 ans.

**Les antécédents personnels ou familiaux** Avoir déjà eu un cancer colorectal ou avoir des antécédents familiaux au premier degré (père, mère, frère, soeur) multiplie le risque par trois. Ceci peut être lié à une prédisposition génétique mais également à l'exposition à une même substance cancérigène ou à des habitudes alimentaires ou un mode de vie commun.



## Facteurs de risques externes

Plusieurs facteurs de risques sont liés au style de vie comme une surcharge pondérale, une consommation excessive de graisses animales et de viande rouge, le tabac, l'alcool et le manque d'activité physique.

### 3.1.3 Traitements

Le traitement pour le cancer colorectal dépend des caractéristiques propres aux patients (santé physique et psychique, âge, etc.) et de celles de son cancer (localisation, évolution, etc.). Traiter le cancer du colon nécessite souvent de recourir à la fois à la chirurgie et à la chimiothérapie, alors que le cancer du rectum peut parfois être traité par radiothérapie.

#### La chirurgie

Le traitement de base du cancer colorectal est la chirurgie. Les médecins pratiquent une intervention appelée colectomie durant laquelle la partie du colon où se situe la tumeur est enlevée et les deux extrémités restantes sont suturées.

#### La chimiothérapie

Lors d'un traitement par chimiothérapie, le patient reçoit une ou plusieurs molécules toxiques pour les cellules cancéreuses (monothérapie ou polythérapie). Les produits les plus utilisés, seuls ou combinés, sont le fluorouracile (5-FU), l'oxaliplatine (Eloxatine®) et l'irinotécan (Campto®).

#### Les thérapies ciblées

Le principal défaut des chimiothérapies usuelles est qu'elles ne s'attaquent pas qu'aux cellules qui sont cancéreuses et peuvent donc engendrer de multiples effets indésirables. Les thérapies ciblées sont moins sujettes à cet inconvénient et elles ont été particulièrement développées dans le cadre du cancer du colon. Néanmoins, elles sont souvent utilisées dans le traitement des cancers colorectaux avancés en association avec de la chimiothérapie.

De telles thérapies vont jouer sur l'angiogénèse en inhibant le développement des vaisseaux sanguins intervenant dans la croissance de la tumeur, et sur les facteurs de croissance, en empêchant la division des cellules et donc la croissance de la tumeur.

Trois molécules peuvent être administrées : le bevacizumab (Avastin®), le cetuximab (Erbix®) et le panitumumab (Vectibix®).

#### La radiothérapie

La radiothérapie est utilisée en traitement local avant l'intervention chirurgicale pour réduire la taille de la tumeur, ou après celle-ci afin de détruire les cellules cancéreuses restantes. Son utilisation reste peu fréquente dans le cas du cancer colorectal.

## 3.2 Protocoles de chimiothérapie

De nombreuses molécules de chimiothérapie peuvent être utilisées, seules ou associées, pour lutter contre le cancer colorectal. Parmi elles, l'irinotécan est particulièrement indiqué dans le cas de cancer colorectal avancé.

---

### 3.2.1 Protocoles incluant de l'irinotécan

#### En combinaison avec le fluorouracile et l'acide folinique

Les patients n'ayant pas été traités avec une chimiothérapie antérieure pour ce stade du cancer peuvent être traités avec de l'irinotécan à une dose de 180 mg/m<sup>2</sup> en perfusion intraveineuse tous les 14 jours, suivie d'une perfusion d'acide folinique et de fluorouracile.

#### En monothérapie

L'irinotécan peut être utilisé en monothérapie, à une dose de 350 mg/m<sup>2</sup> administrée en perfusion intraveineuse tous les 21 jours après abandon d'un traitement à base de fluorouracile. Le patient peut aussi être traité tous les 14 jours dans certains cas.

#### En combinaison avec le cetuximab

Le cetuximab peut être combiné à l'irinotécan lorsque le patient est traité pour un cancer colorectal métastatique avec gène KRAS de type sauvage, qu'il n'a pas encore été traité pour une maladie métastatique ou suite à l'abandon d'une chimiothérapie utilisant de l'irinotécan.

#### En combinaison avec le fluorouracile, l'acide folinique et le bevacizumab

Dans le traitement de première ligne des patients présentant un carcinome métastatique du colon ou du rectum, l'irinotécan est parfois utilisé en combinaison avec le fluorouracile, l'acide folinique et le bevacizumab.

#### En combinaison avec la capecitabine avec ou sans bevacizumab

Les patients présentant un cancer colorectal métastatique peuvent recevoir de l'irinotécan en association avec de la capecitabine en traitement de première ligne.

### 3.2.2 Adaptation de la posologie

#### Allongement du cycle de traitement

Il est recommandé de reporter la date d'administration de l'irinotécan d'une ou deux semaines tant que toutes les toxicités n'ont pas pour grade 0 ou 1, et que le patient subit une diarrhée liée à la chimiothérapie.

#### Réduction de doses

Il est recommandé de réduire de 15 à 20 % les doses d'irinotécan et/ou de fluorouracile lors de l'apparition de :

- toxicité hématologique : neutropénie grade 4, neutropénie fébrile (neutropénie grade 3-4 et fièvre grade 2-4), thrombopénie et leucopénie (grade 4) ;
- toxicité non hématologique (grade 3-4).

### 3.2.3 Toxicités

Les toxicités intervenant dans plus de 10% des cas et nécessitant une réduction de la dose d'irinotécan sont les diarrhées tardives et les troubles hématologiques (neutropénie, thrombopénie et anémie).

Les toxicités du Tableau 3.1 ont par exemple été décelées chez 145 patients ayant un reçu une association d'irinotécan, de fluorouracile et d'acide folinique tous les 14 jours, avec une dose recommandée

d'irinotécan de 180 mg/m<sup>2</sup> (<http://base-donnees-publique.medicaments.gouv.fr/affichageDoc.php?specid=60306813&typedoc=R>).

Tableau 3.1 – Effets indésirables rapportés avec l'irinotécan en traitement combiné (protocole de 180 mg/m<sup>2</sup> toutes les 2 semaines).

Classe de système d'organe	Classe de fréquence	Terme préféré
Infections et infestations	Fréquent	Infection
Affections du sang et du système lymphatique	Très fréquent Très fréquent Très fréquent Fréquent	Thrombopénie Neutropénie Anémie Neutropénie fébrile
Troubles du métabolisme et de la nutrition	Très fréquent	Perte d'appétit
Affections du système nerveux	Très fréquent	Syndrome cholinergique
Affections gastro-intestinales	Très fréquent Très fréquent Très fréquent Fréquent Fréquent	Diarrhées Vomissements Nausées Douleurs abdominales Constipation
Affections de la peau et du tissu sous-cutané	Très fréquent	Alopécie (réversible)
Troubles généraux et anomalies au site d'administration	Très fréquent Très fréquent Fréquent	Inflammation des muqueuses Asthénie Hyperthermie
Investigations	Très fréquent Très fréquent Très fréquent	Élévation des transaminases (SGPT et SGOT) Augmentation de la bilirubine Augmentation des phosphatases alcalines sanguines

### 3.3 Données disponibles

Afin d'assurer le report complet de nos données de santé de routine, nous avons suivi les indications RECORD [7].

#### 3.3.1 Patients inclus

Les données de chimiothérapies des patients sont issues de dossiers de santé informatisés et stockées sur la plateforme i2b2 dans un entrepôt de données cliniques [116, 56]. Ces entrepôts de données ont été construits afin de faciliter la réutilisation des données de prise en charge médicale pour la recherche. Afin de pouvoir analyser un ensemble de patients de taille raisonnablement grand tout en assurant un minimum d'uniformité entre les traitements reçus, nous avons extrait de l'entrepôt i2b2 de l'Hôpital Européen Georges Pompidou (HEGP) des patients traités pour un cancer colorectal métastatique et ayant reçu au moins une fois une combinaison de molécules incluant l'irinotécan à une dose théorique de 180 mg/m<sup>2</sup> et avec un cycle théorique de quatorze jours entre le 1er Juillet 2003 et le 10 Novembre 2017.

#### 3.3.2 Variables extraites

Les protocoles de chimiothérapie et les prescriptions sont extraits du logiciel Chimio® qui est utilisé par les médecins pour leurs prescriptions. Ce logiciel gère les dates d'administration de chacun des cycles, les doses des différentes molécules, les réductions de doses, etc. Certaines données cliniques sont aussi stockées par les cliniciens dans Chimio® comme l'indice de masse corporelle, la surface corporelle ou la clairance de la créatinine par exemple.

Les données de l'application Chimio® ont été importées dans l'entrepôt de données i2b2.

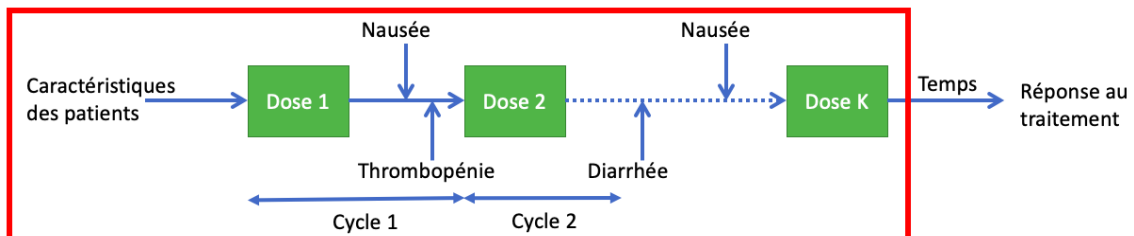
Finalement, 1,4 millions d’observations ont été importées dans les dossiers patients informatisés de Chimio® dans i2b2, impliquant 14000 patients différents. 1300 concepts i2b2 ont été créés pour stocker ces observations.

Concernant les données cliniques associées à chaque cycle de chimiothérapie, un questionnaire spécifique a été mis en place en 2012 à l’HEGP pour collecter les items cliniques nécessaires au suivi du patient et aux ajustements de dose (cf. Annexe A). En particulier, il contient les toxicités de chimiothérapie à chaque cycle graduées par organe et sévérité selon la nomenclature CTCAE. Nous avons extrait des toxicités ainsi que d’autres variables pertinentes (âge, score OMS, bilirubine, perte de poids, ligne de traitement).

Nous avons dérivé des données extraites les variables pertinentes suivantes :

- Âge inférieur ou supérieur à 80 ans ;
- Perte de poids inférieur ou supérieur à 10 % (depuis le début du traitement) ;
- Bilirubine catégorisée entre bilirubine inférieure à  $35 \mu\text{mol/L}$ , entre 35 et  $50 \mu\text{mol/L}$  et supérieure à  $50 \mu\text{mol/L}$  ;
- Ligne de traitement inférieure à 3, égale à 3 ou supérieure à 3 ;
- Score OMS et toxicités gradués par des entiers de 0 (pas d’effets indésirables) à 3 ou 4 (grade le plus sévère).

Toutes ces variables sont donc connues pour chaque patient et chacun de ses cycles, il s’agit donc de données répétées. Dans notre cas, le traitement est donné le jour 1 et le prochain cycle commence 14 jours après. Des effets indésirables peuvent apparaître sur le cycle, tel que le décrit la Figure 3.1. Le jeu de données a été construit de manière à étudier l’impact des variables sur la décision du médecin d’adapter la dose, donc chaque ligne du jeu de données correspond à un cycle d’un patient donné et contient la dose donnée au patient au début du cycle et les variables observées entre le début du cycle précédent et le cycle actuel.

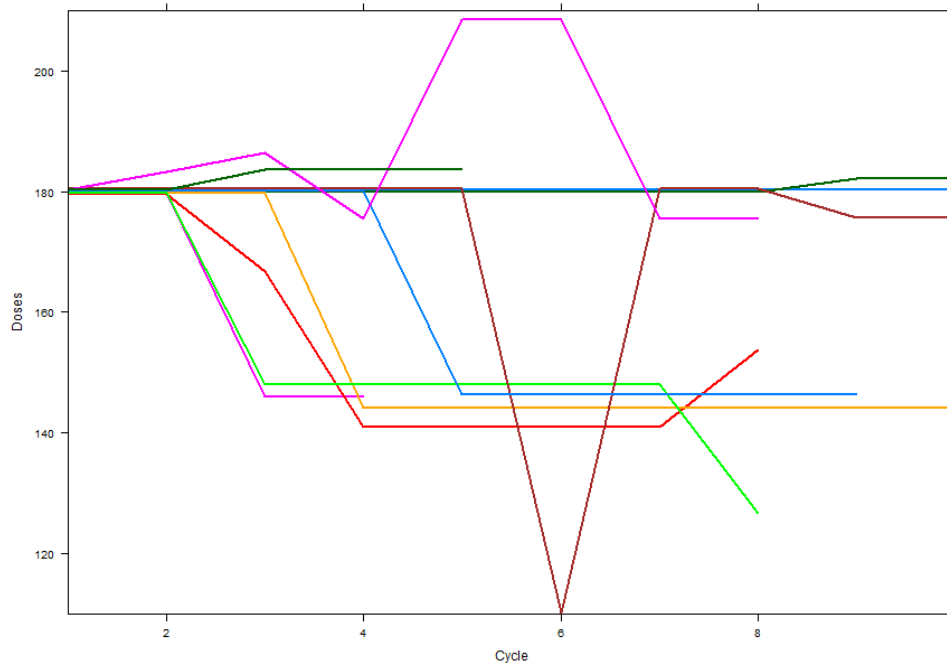


Graphique 3.1 – Exemple de l’ajustement de dose pour un patient en fonction de ses caractéristiques et des toxicités.

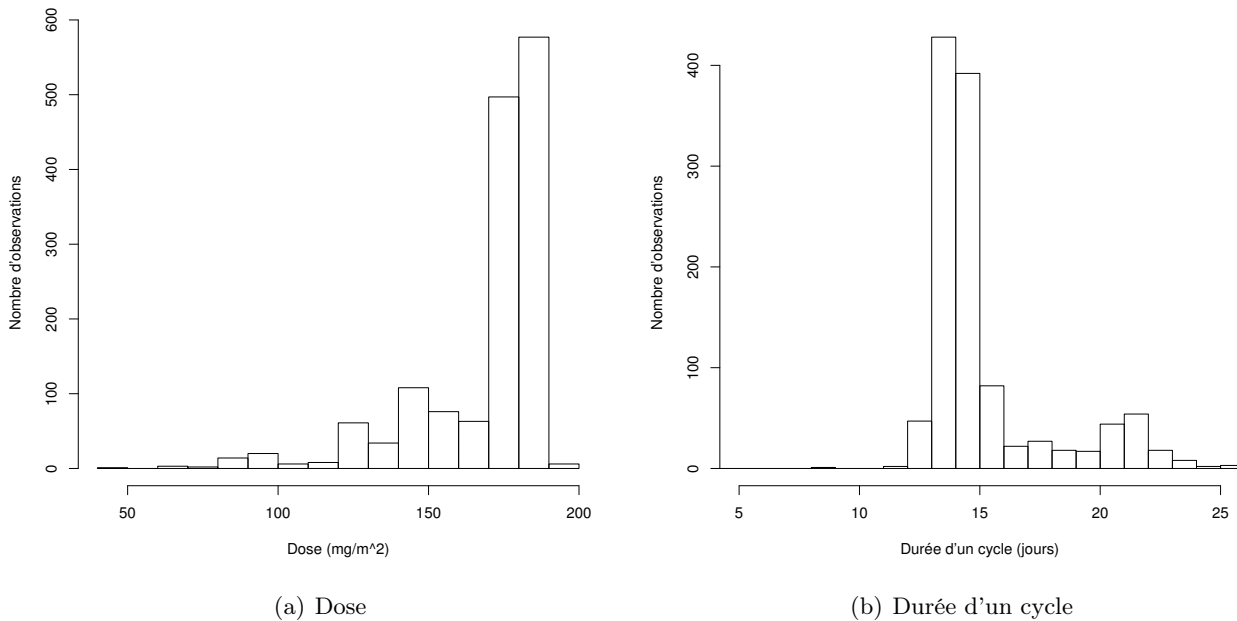
### 3.3.3 Analyse descriptive

Notre base de données compte 678 patients. Les Figures 3.2 et 3.3(a) présentent respectivement les trajectoires individuelles des doses pour plusieurs patients sur plusieurs cycles et la distribution de la dose sur l’ensemble des cycles. Ainsi, bien qu’une majorité de doses soit proche de la dose de référence de  $180 \text{ mg/m}^2$ , beaucoup de patients reçoivent aussi des doses réduites au cours du temps. Au lieu de réduire la dose, les cliniciens choisissent parfois aussi d’allonger la durée entre deux cycles comme le montre la figure 3.3(b). Le Tableau 3.2 donne les effectifs et pourcentages associés à chaque classe de chaque variable. Par exemple, 44 observations (3%) sont associées à une perte de poids d’au moins 10 % contre 1433 (97 %) qui ne le sont pas. Les Figures 3.4 et 3.5 présentent les distributions respectives des caractéristiques des patients et de leurs toxicités. Très peu d’observations de toxicités de grade 3 ou 4 ont été recueillies. La base de données compte également énormément de données manquantes, ce qui rend difficile l’utilisation des méthodes de gestion de données manquantes usuelles. En effet, en pratique clinique, ce sont souvent de jeunes internes qui remplissent les questionnaires des patients, et ceux-là

sont, par manque de temps, souvent contraints de ne compléter qu'une petite partie des informations demandées dans le questionnaire. D'où, toujours, cette idée qu'il est important de sélectionner un ensemble optimal et limité de variables pour construire un outil d'aide à la décision.



Graphique 3.2 – Exemple de l'évolution de la dose sur plusieurs cycles pour plusieurs patients.



Graphique 3.3 – Distributions de la dose et de la durée des cycles.

Tableau 3.2 – Statistiques descriptives sur l'ensemble des cycles. Les pourcentages sont donnés entre parenthèses.

Variables	NA	0	1	2	3	4
Âge $\geq 80$ ans	-	1384 (94)	92 (6)	-	-	-
Perte de poids $> 10\%$	-	1433 (97)	44 (3)	-	-	-
Score OMS	656 (44)	160 (20)	493 (60)	131 (16)	14 (2)	22 (3)
Bilirubine $>35, >50 \mu\text{mol/L}$	661 (45)	803 (99)	2 (0)	10 (1)	-	-
Ligne de traitement 3, $> 3$	740 (50)	637 (87)	54 (7)	45 (6)	-	-
<b>Grades de toxicité</b>						
Vomissements	680 (46)	707 (89)	69 (9)	16 (2)	4 (1)	-
Nausées	671 (45)	434 (54)	283 (35)	80 (10)	8 (1)	-
Diarrhées	663 (45)	460 (57)	280 (34)	65 (8)	8 (1)	-
Asthénie	681 (46)	129 (16)	463 (58)	178 (22)	25 (3)	-
Neutropénie	625 (42)	625 (73)	128 (15)	79 (9)	13 (2)	6 (1)
Thrombopénie	616 (42)	714 (83)	134 (16)	8 (1)	2 (0)	2 (0)
Anémie	607 (41)	344 (40)	51 (6)	1 (0)	1 (0)	-

### 3.4 Élicitation des poids de sévérité

Pour chaque variable  $j \in \{1, \dots, J\}$  et chacune de ses valeurs possibles  $c \in \{1, \dots, C_j\}$ , les médecins devaient définir des poids de pertinence clinique  $\mathbf{w}_j = (w_{j,1}, \dots, w_{j,C_j})$ . Ces poids montrent quelles variables ont le plus d'importance dans la décision d'adapter la dose selon l'expertise de chaque clinicien. Ces poids doivent être définis tels que :  $0 = w_{j,0} \leq w_{j,1} \leq \dots \leq w_{j,C_j} \leq W_{max} = 100$ , ce qui signifie que des variables associées à des poids faibles auront moins d'influence dans la décision de réduire la dose que des variables liées à des poids importants.

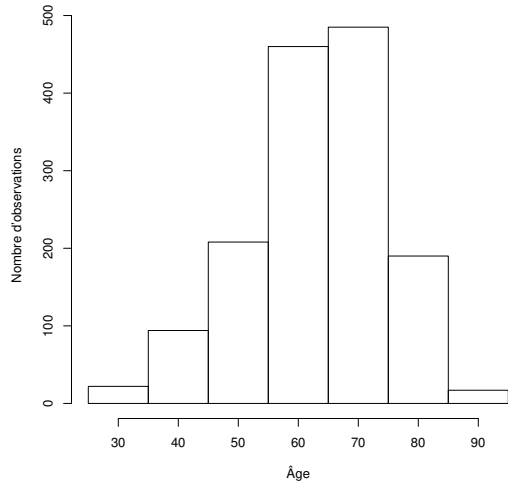
Pour chaque variable  $j \in \{1, \dots, J\}$ ,  $w_{j,0} = 0$  est le poids associé au niveau de référence  $x'_{j,0}$ . De manière similaire, pour  $c \in \{1, \dots, C_j\}$ ,  $w_{j,c}$  est le poids associé quand la variable  $x_{i,k,j}$ ,  $i \in \{1, \dots, n\}$ ,  $k \in \{1, \dots, K_i\}$  prend la valeur  $x'_{j,c}$ . Les valeurs minimale et maximale possibles pour les poids ont été choisies à 0 (aucune gravité) et  $W_{max} = 100$  (le plus grave possible) car cette échelle est facilement compréhensible par les oncologues. Par exemple, le vecteur  $\mathbf{w}_7 = (0, 20, 80)$  est le vecteur des poids associés à la 7ème variable et à chacun de ses  $C_j = 3$  niveaux (le niveau de référence est exclu). Par analogie avec  $\mathbf{z}$ ,  $\mathbf{w}^*$  est défini comme le vecteur de taille  $L$  de tous les poids de pertinence clinique. Toutes les notations des poids sont issues du travail de Bekele et Thall (2004) dans lequel des poids élicités de sévérité associés à des toxicités ont été utilisés pour construire un score appelé "*total toxicity burden*" [6].

Pour notre projet, les cancérologues nous ont aidés à construire un questionnaire en nous donnant les variables et les seuils qu'ils utilisent pour prendre leur décision. Six cliniciens ont individuellement complété le questionnaire en ligne (cf. Annexe B). Chaque clinicien a attribué un poids numérique à chaque grade de chaque type de toxicité et chaque niveau de chaque variable. Lorsque les médecins ont été interrogés sur d'éventuelles variables absentes du questionnaire et susceptibles d'influencer les doses, seules deux variables supplémentaires ont été mentionnées : l'objectif thérapeutique (deux cliniciens) et l'obstruction intestinale (un clinicien).

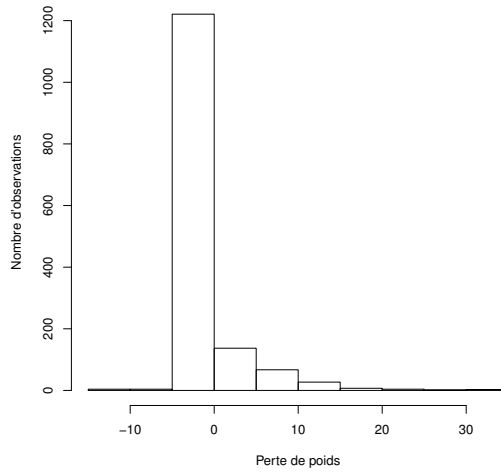
Le Tableau 3.3 donne les poids de pertinence clinique élicités par chaque clinicien pour chaque variable. Par exemple,  $\{0, 20, 80\}$  dans la première colonne (clinicien 1), sur la ligne "nausée", correspond aux poids de pertinence clinique donnés par le médecin 1 pour les grades  $\{1, 2, 3\}$ . Le tableau suggère que le clinicien 2 assigne un poids de 20 à une perte de poids de plus de 10 % dans sa décision de réduire la dose, alors que le clinicien 4 lui attribue un poids de 80.

Tableau 3.3 – Poids de pertinence clinique élicités par chaque clinicien pour chacune des variables.

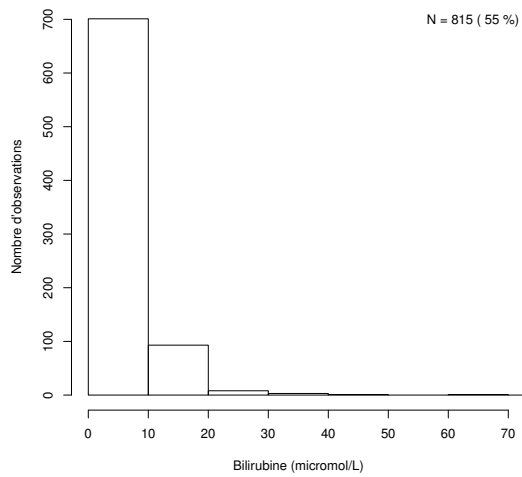
Variables	Médecins											
	1				2				3			
Age $\geq$ 80 ans	100	-	-	-	60	-	-	-	100	-	-	-
Perte de poids $>$ 10%	50	-	-	-	20	-	-	-	50	-	-	-
Score OMS (1,2,3,4)	0	20	20	20	0	0	40	100	0	0	80	100
Bilirubine $>$ 35, $>$ 50 $\mu\text{mol/L}$	100	100	-	-	40	80	-	-	100	100	-	-
Ligne de traitement 3, $>$ 3	30	50	-	-	0	0	-	-	0	0	-	-
<b>Toxicités de grades 1, 2, 3, 4</b>												
Vomissements	0	20	80	90	0	30	70	100	0	10	10	10
Nausées	0	20	80	-	0	10	50	-	0	10	10	-
Diarrhées	0	40	80	100	0	20	50	100	0	50	80	100
Asthénie	10	50	100	-	10	10	40	-	0	0	70	-
Neutropénie	0	70	100	100	0	0	30	50	0	0	50	50
Thrombopénie	40	100	100	100	0	0	20	30	0	0	50	50
Anémie	0	50	80	100	0	0	20	30	0	0	0	0
Variables	Médecins											
	4				5				6			
Age $\geq$ 80 ans	80	-	-	-	100	-	-	-	100	-	-	-
Perte de poids $>$ 10%	80	-	-	-	50	-	-	-	60	-	-	-
Score OMS (1,2,3,4)	0	20	80	100	0	30	100	100	0	70	100	100
Bilirubine $>$ 35, $>$ 50 $\mu\text{mol/L}$	20	80	-	-	100	100	-	-	100	100	-	-
Ligne de traitement 3, $>$ 3	0	0	-	-	0	0	-	-	0	50	-	-
<b>Toxicités de grades 1, 2, 3, 4</b>												
Vomissements	10	20	80	100	0	30	70	100	0	0	70	100
Nausées	10	30	80	-	0	10	40	-	0	0	30	-
Diarrhées	0	20	70	90	0	20	50	100	0	50	100	100
Asthénie	10	50	70	-	0	20	50	-	0	40	80	-
Neutropénie	0	20	70	80	0	0	20	100	0	0	0	50
Thrombopénie	0	50	80	100	0	0	20	70	0	0	0	0
Anémie	0	20	50	70	0	0	30	80	0	0	0	0



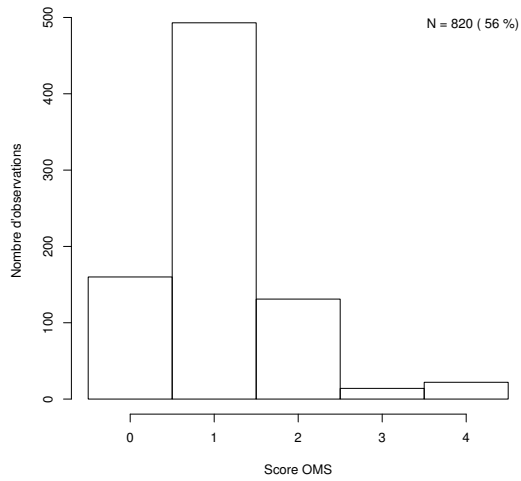
(a) Âge



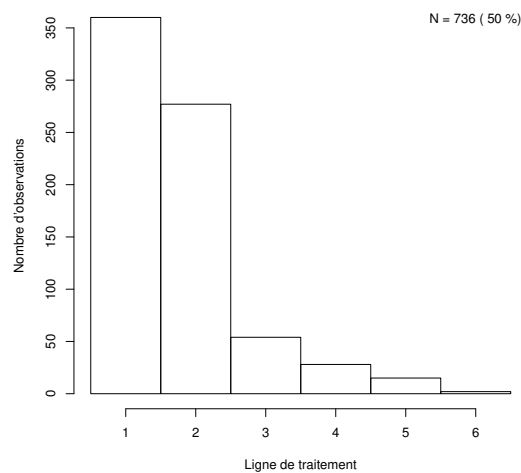
(b) Perte de poids



(c) Bilirubine



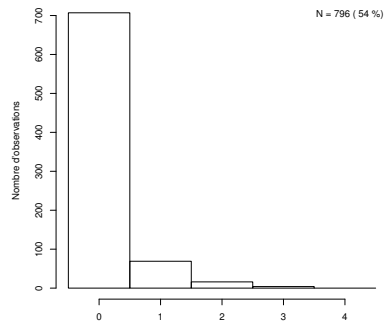
(d) Score OMS



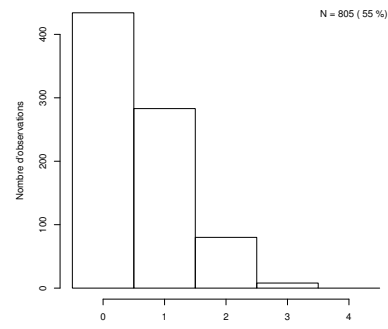
(e) Ligne de traitement

Graphique 3.4 – Distributions de l'âge, de la perte de poids, de la bilirubine, du score OMS et de la ligne de traitement sur l'ensemble des cycles.

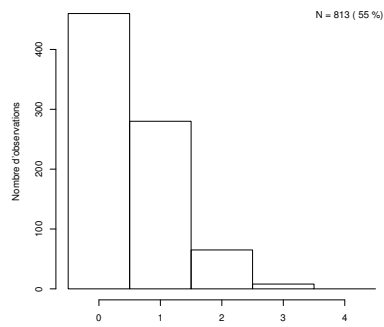




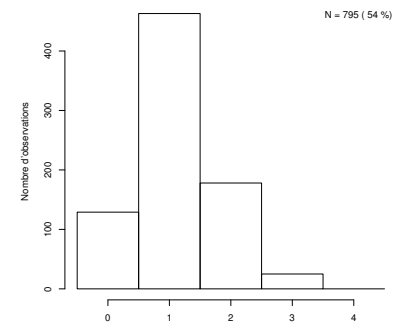
(a) Vomissements



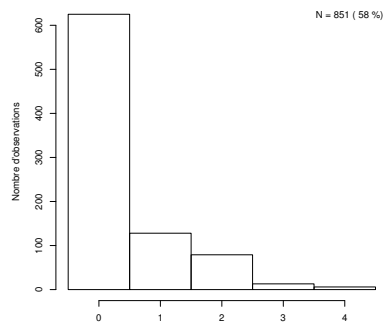
(b) Nausées



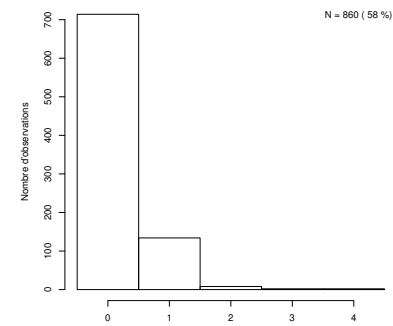
(c) Diarrhées



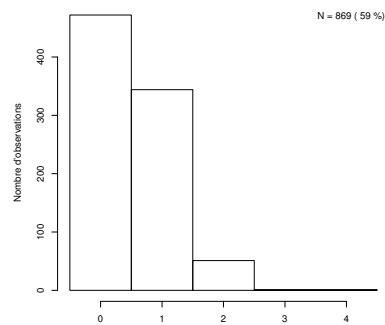
(d) Asthénie



(e) Neutropénie



(f) Thrombopénie



(g) Anémie

Graphique 3.5 – Distribution des toxicités sur l'ensemble des cycles.

# Méthode de sélection de variables Bayésienne utilisant des poids de pertinence clinique dans le cas de petits échantillons - Application au cancer du colon

**Contexte :** La réutilisation des dossiers de santé électroniques des hôpitaux peut faciliter une meilleure compréhension de la prise de décision médicale associée aux décisions de traitement séquentiel adaptatives. Cependant, modéliser les décisions médicales derrière les choix de traitement en fonction des caractéristiques du patient soulève des problèmes méthodologiques. En cancérologie, dans la pratique clinique, lorsque les cliniciens doivent adapter les doses de chimiothérapie, ils vont implicitement attribuer plus ou moins de poids aux caractéristiques du patient (âge, perte de poids,...), aux effets indésirables antérieurs et à l'évolution du cancer dans leur décision d'adapter les doses. Face aux petits échantillons, les méthodes de sélection Bayésienne permettent de sélectionner les variables réellement utilisées en pratique clinique et d'intégrer l'information provenant des experts dans les distributions a priori. L'originalité de ce travail est de combiner l'élicitation des soins de routine avec les données observées. Dans le contexte des données de vie réelle impliquant des adaptations de doses d'irinotécan répétées pour le cancer métastatique colorectal, nous proposons une modification de la méthode SSVS (*Stochastic Search Variable Selection*) basée sur des poids de pertinence clinique élicités qui permettent d'incorporer l'expertise médicale dans les distributions a priori.

**Méthodes :** Notre travail comporte deux étapes principales : l'élicitation des poids de pertinence clinique et la construction des distributions a priori. Les poids de pertinence clinique reflètent l'importance qu'un médecin donne à chaque valeur de chaque variable utilisée dans la décision médicale. Dans notre situation, les oncologues nous ont aidé à construire un questionnaire en choisissant les variables et les seuils qu'ils utilisent pour la prise de décision. Puis, quatre cliniciens experts ont complété séparément le questionnaire en ligne. Chaque clinicien a spécifié un poids de pertinence clinique numérique pour chaque grade de chaque type de toxicité et chaque niveau de chaque variable, dans l'intervalle  $[0,100]$ , où 0 correspond à une absence de gravité et 100 au plus haut niveau de sévérité. Le Tableau 3.3 donne les poids de pertinence clinique élicités pour chaque expert et chaque variable. Par exemple, la ligne "vomissements" du Tableau 3.3 renvoie aux poids de pertinence clinique donnés par le clinicien 1. Le tableau suggère que si des vomissements de grade 2 sont observés, le clinicien 1 réduira la dose dans 20 % des cas.

Ces poids élicités ont été utilisés pour construire les distributions a priori Beta des indicatrices de sélection de variables de la méthode SSVS modifiée, appelée *Weight-Based SSVS* (WBS). Afin d'identifier

les variables qui ont une influence sur l'ajustement de dose, la relation entre les ajustements répétés de doses d'irinotécan en fonction des caractéristiques des patients et des toxicités, dans le cancer colorectal métastatique, a été modélisée par un modèle linéaire à effets mixtes.

**Simulation :** Nous avons analysé les performances du modèle WBS comparées à celles des modèles LASSO et SSVS par le biais d'une étude de simulation approfondie. Pour quatre scénarios différents, nous avons simulé 1000 jeux de données. Chaque jeu de donnée était constitué d'un échantillon de patients traités pour un cancer colorectal métastatique par irinotécan et a été simulé suivant les distributions de notre jeu de données réelles observées.

**Résultats :** Comparée à la méthode SSVS usuelle, la méthode WBS a montré de meilleures performances quelque soit le critère de performance considéré et a produit les plus faibles taux de faux positifs et de faux négatifs. Comme attendu, la performance s'améliore avec l'accroissement de la taille de l'échantillon. La performance du modèle WBS dépend non seulement des poids associés aux variables par les cliniciens mais aussi de la somme de ces poids qui doit donc être scrupuleusement calibrée pour sélectionner un nombre approprié de variables. Le modèle LASSO a montré de mauvaises performances comparé aux modèles SSVS et WBS confirmant l'avantage de l'utilisation de l'approche Bayésienne sur l'approche fréquentiste dans le cadre des petits échantillons.

Nous avons également appliqué notre méthode WBS à notre cas d'usage. Concernant l'élicitation des poids de pertinence clinique, notre cas d'usage a suggéré que les cliniciens pouvaient ne pas avoir une perception correcte de leur processus de décision réelle concernant l'ajustement de dose. En effet, la plupart des personnes sont influencées par des facteurs cognitifs, psychologiques et émotionnels qui peuvent les empêcher, lors de prise de décision, de choisir la meilleure option possible d'un point de vue rationnel [41]. Par exemple, en pratique clinique, si un médecin constate des effets indésirables graves chez un patient un jour donné, ses émotions et son expérience personnelle peuvent l'amener à réduire la dose du prochain patient, même si le bon sens voudrait que la dose soit maintenue.

**Conclusion :** Ce travail propose une méthodologie Bayésienne novatrice permettant d'incorporer des poids de pertinence clinique élicités par des experts dans une analyse de données. À notre connaissance, la littérature portant sur les méthodes de sélection de variables incorporant des distributions a priori informatives est extrêmement limitée dans le cas de variables provenant de la pratique quotidienne, comme des dossiers de santé électroniques des hôpitaux. La modélisation de la prise de décision médicale peut être vue comme la première étape dans la modélisation de la relation complexe entre variables, décisions de réduction de doses et survie. Notre méthodologie pourrait être étendue pour se focaliser sur les combinaisons de doses permettant d'optimiser la survie chez les patients davantage sujets aux effets indésirables. Notre méthode pourrait aussi être appliquée dans d'autres situations où nous souhaitons prendre en compte l'opinion d'experts pour sélectionner des variables spécifiques dans le cas de petits échantillons.

## RESEARCH ARTICLE

# Bayesian variable selection based on clinical relevance weights in small sample studies—Application to colon cancer

Sandrine Boulet<sup>1</sup>  | Moreno Ursino<sup>1</sup>  | Peter Thall<sup>2</sup> | Anne-Sophie Jannot<sup>1,3</sup> | Sarah Zohar<sup>1</sup>

<sup>1</sup>INSERM U1138, Team 22, Centre de Recherche des Cordeliers, University Paris Descartes, University Pierre et Marie Curie, Paris, France

<sup>2</sup>Department of Biostatistics, MD Anderson Cancer Center, Houston, Texas

<sup>3</sup>Department of Statistics, Computing and Public Health, Hôpital Européen Georges-Pompidou, AP-HP, Paris, France

**Correspondence**

Sandrine Boulet, INSERM U1138, Team 22, Centre de Recherche des Cordeliers, University Paris Descartes, University Pierre et Marie Curie, Paris, France.  
Email: sandrine.boulet@crc.jussieu.fr

**Funding information**

French National Cancer Institut (INCa), Grant/Award Number: INCA\_10801 and INCA\_9539

Using clinical data to model the medical decisions behind sequential treatment actions raises methodological challenges. Physicians often have access to many covariates that may be used when making sequential treatment decisions for individual patients. Statistical variable selection methods may help finding which of these variables are used for this decision in everyday practice. When the sample size is not large, Bayesian variable selection methods can address this setting and allow for expert information to be incorporated into prior distributions. Motivated by clinical practice data involving repeated dose adaptation for Irinotecan in colorectal metastatic cancer, we propose a modification of the stochastic search variable selection (SSVS) method, which we call weight-based SSVS (WBS). We use clinical relevance weights elicited from physician experts to construct prior distributions, with the goal to identify the most influential toxicities and other covariates used for dose adjustment. We evaluate and compare the WBS model performance to the Lasso and SSVS through an extensive simulation study. The simulations show that WBS has better performance and lower rates of false positives and false negatives than the other methods but depends strongly on the covariate weights.

**KEYWORDS**

clinical relevance weights elicitation, informative priors, repeated measures, stochastic search variable selection

## 1 | INTRODUCTION

Recent innovations in the accessibility and availability of hospital electronic health records (EHRs), where comprehensive information about patients are gathered, have made this source of structured data accessible for statistical analysis and medical research.<sup>1,2</sup> Reuse of such records facilitates a better understanding of medical decision making associated with sequential treatment decisions. Modeling the medical decisions behind treatment actions as a function of patient characteristics, including previous treatments or doses and clinical outcomes, raises methodological challenges. Statistical models must account for many possible covariates, comorbidities, and clinical events that may explain dose modifications or sequences of treatments for individual patients.

**Abbreviations:** SSVS, stochastic search variable selection; RMSE, root mean square error; FPR, false positive rate; FNR, false negative rate

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2019 The Authors *Statistics in Medicine* Published by John Wiley & Sons Ltd

General treatment recommendations often are based on clinical trial results, in which patients are a selected subset of the actual treated population. Inclusion and exclusion criteria of clinical trials can be so strict that older, pediatric, or patients with particular comorbidities may be excluded. As a result, general recommendations may be unsuitable for these patients, which forces practicing physicians to modify treatment decisions based on each patient's characteristics and history. In medical practice, at each patient visit, the physician must choose a new treatment or modify the dose of an ongoing treatment based on their experience, outside recommendations, and the patient's current history, including covariates, previous treatments, and clinical outcomes. Modeling such medical decision making based on observed toxicities and covariates can be difficult, particularly because of the multidimensionality of clinical toxicity observations, classified by organ and severity ([https://ctep.cancer.gov/protocolDevelopment/electronic\\_applications/ctc.htm](https://ctep.cancer.gov/protocolDevelopment/electronic_applications/ctc.htm)). Physicians typically prioritize certain covariates over others and may use their own subjective clinical relevance weights when making treatment decisions. Accounting for clinical relevance weights may be highly informative when analyzing EHR data to estimate optimal treatment sequences or dose modifications.

In oncology, many innovative molecules and biomarkers (genomic, clinical, or physiological) have been recently proposed, which has led to consideration of many small subpopulations of diseases previously assumed to be homogeneous.<sup>3</sup> In such small subpopulations, identifying causes of toxicities and resulting treatment modifications in everyday health care may be quite complicated. With small samples, fitted multivariate models may have large variances of estimated parameters, limiting inferential reliability. Variable selection methods in such settings may mitigate this difficulty by keeping only covariates having the strongest effects. Most of these methods are frequentist, using penalization criteria, such as the bridge, Lasso, SCAD, LARS, elastic net, or OSCAR regressions.<sup>4-8</sup> One of the most popular methods is the Lasso, which performs variable selection by applying an L1-penalty to least squares. The Lasso shrinks some coefficients to zero while selecting others. However, these variable selection methods have been validated and associated with good convergence properties for large samples. In our setting, given the small sample sizes and large number of variables considered, these methods often fail to provide reliable results.

Bayesian regression is a reasonable alternative approach when working with small samples. In the Bayesian framework, variable selection may be performed by applying the so-called "spike and slab" priors to identify and estimate the posteriors of nonzero regression parameters<sup>9</sup> while deleting variables that do not appear to predict outcome. Mitchell and Beauchamp assumed the prior distribution of each regression coefficient to be a two-component mixture of a point mass at 0 (the spike) and a uniform distribution on a finite interval (the slab).<sup>10</sup> These two components quantify prior belief about whether a covariate effect exists and, if it does, the effect's magnitude. This method has been extended in a variety of ways, focusing on specific formulations of priors.<sup>11-15</sup> A very useful extension is the stochastic search variable selection (SSVS) method, in which the prior of each parameter is assumed to be a mixture of two normal distributions centered at zero but with very different variances,<sup>11</sup> one large and the other small. Other methods specify continuous prior distributions that approximate the "spike and slab" shape while shrinking the estimates toward zero.<sup>16,17</sup> Park and Casella extended the frequentist Lasso by assuming a double exponential distribution for the regression parameters.<sup>17</sup> For a review of Bayesian variable selection methods, see the work of O'Hara and Sillanpää.<sup>9</sup>

A major advantage of Bayesian inference is that external information can be incorporated formally into prior distributions. Many Bayesian variable selection studies use priors that do not specify a preference for any variable but may, for instance, show a preference for sparse models. Several Bayesian variable selection studies have proposed simple approaches for incorporating prior knowledge, by independently assigning each variable a prior probability of being included in the model.<sup>18</sup> For example, Kitchen et al provide a method for prior parameter specification using the scientific literature for regression problems with many predictors in which the priors include both shrinkage and variable selection components, extending the model of Kuo and Mallick.<sup>12</sup> The authors posit this approach as a solution to the phenotype-genotype problem. They construct a so-called "voting" prior by assigning a count to each codon position based on the number of times that it is identified as important in the literature.<sup>19</sup> In such a study, the same amount of information is available for each biomarker. However, in other settings with many covariates, such as clinical variables, the variable selection literature is limited.

We are motivated by clinical practice data in which dose modifications were made over multiple cycles for patients being treated with Irinotecan for colorectal cancer. We use elicited clinical relevance weights to reflect the medical experience of practicing physicians.<sup>20</sup> When a physician makes a decision regarding a patient's treatment or dose modification, certain patient characteristics and previous clinical outcomes are given more weight than others. This practice implies that the decision is influenced by the physician's personal experience and knowledge regarding the clinical relevance of certain covariates. To analyze our motivating data set, we elicited expert relevance weights for patient covariates, including toxicities, from each of several physicians and used these weights to construct prior distributions.

In the colorectal cancer setting, biomarkers that explain certain toxicities have recently been identified. For example, activating mutations of the KRAS oncogene have been shown to be predictive of resistance of anti-EGFR agents, such as Cetuximab, which is sometimes used with Irinotecan in treating metastatic colorectal cancer.<sup>21</sup> Another example is the gene UGT1A1, which is strongly associated with bilirubin levels, for which a variant is linked to an increased risk of Irinotecan toxicity. This finding suggests that patient subgroups defined by these variables should be treated differently. The most common adverse events associated with Irinotecan include vomiting, nausea, diarrhea, asthenia, neutropenia, and anemia. All of these toxicities occur at various severity levels, classified as grade 0, 1, 2, 3, or 4. Generally, the dose for each patient is calculated according to his/her body surface area. Toxicities often require decreasing the patient's dose, extending the delay between two dose administration intervals (or cycles), or temporarily discontinuing treatment. In addition, age, performance status, bilirubin, genetic polymorphism, and drug administration schedule have been shown to be related to Irinotecan toxicity.<sup>22</sup> While standard protocols for dose adjustment in chemotherapy for adverse events exist, these recommendations do not account for possible associations between the covariates noted above. Thus, in clinical practice, physicians adapt doses and schedules during successive cycles of treatment by accounting for all of the patient's characteristics and treatment history, including doses and outcomes. As a result, each patient's dose regimen over multiple cycles often differs from what is recommended in standard protocols.

To analyze the Irinotecan data, we proceeded in four steps: (1) First, we asked each of four physicians who are experts in treating colorectal cancer to specify numerical clinical relevance weights to reflect their beliefs about the importance of each variable in their decision making. When a covariate takes on a specific value, physicians will reduce the dose with a probability that depends on the physician's clinical relevance weight associated with this covariate. Each physician provided their clinical relevance weight associated with each grade of each toxicity type and each level of each covariate. (2) Next, we constructed a modification of the SSVS method, which we call weight-based SSVS (WBS), to identify covariates that influence dose adjustment. We used the elicited clinical relevance weights to compute the hyperparameters of the prior distributions of the variables' inclusion indicators to provide a basis for the WBS. (3) Third, we compared the influence of the physicians' clinical relevance weights using our proposed method with two other methods, the Lasso and conventional SSVS. We also studied the effect of sample size on our method. (4) Finally, we applied our WBS method to analyze the Irinotecan data set.

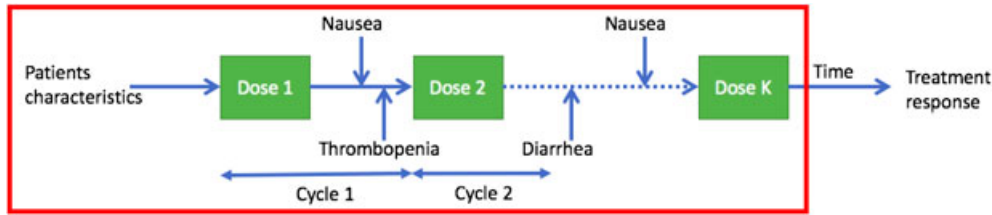
The remainder of this paper is organized as follows. Methods and criteria for model comparison are discussed in Section 2. In Section 3, we describe a simulation study design to assess our methodology, and we present the simulation results in Section 4. We present the data analysis of our case study in Section 5 and close with a discussion in Section 6.

## 2 | METHODS

Our motivating data set is made of one continuous dependent variable, the dose of Irinotecan ( $\text{mg}/\text{m}^2$ ), and twelve covariates: age  $> 80$  years, weight loss  $> 10\%$  since the beginning of treatment, World Health Organization (WHO) score, bilirubin  $> 35 \mu\text{mol}/\text{L}$ , treatment line  $\geq 3$  (that is, the patient previously received more than three other cancer treatments) and toxicities associated with Irinotecan treatment (vomiting, nausea, diarrhea, asthenia, neutropenia, thrombopenia, and anemia). All of these variables are known for each patient and each of his/her cycles, so it is repeated measures data. In cancer, a cycle is a course of treatment repeated on a regular schedule with periods of rest in between. In our setting, the treatment is given on day 1 and the next cycle begins 14 days later. Thus, a new cycle starts when the patient receives a new dose. Side effects can appear during the cycle, as described in Figure 1. The data set was built to study the impact of covariates on clinicians' decisions to adapt the doses, therefore each line of the data set corresponds to one cycle of a given patient and contains the dose given to the patient at the beginning of the cycle and the observed covariates between the start of the previous cycle and the present cycle.

The following notation is introduced for the general problem of variable selection using clinical relevance weights for longitudinal data. Let  $n$  be the number of subjects,  $J$  the number of covariates, and  $K$  the number of cycles. For patient  $i \in \{1, \dots, n\}$ , let  $\mathbf{x}_{i,k}$  be the  $J$ -vector of covariates available at the start of cycle  $k \in \{1, \dots, K\}$ . In our application, certain variables, such as age and treatment, are assumed to be fixed in time, while other patient characteristics and toxicities associated with adverse events and treatment responses may change over time. For  $j \in \{1, \dots, J\}$ , let  $x_{i,k,j}$  take values in  $\{x'_{j,0}, x'_{j,1}, \dots, x'_{j,C_j}\}$ , where  $x'_{j,c}$  is the  $c$ th most severe level,  $c = 0, \dots, C_j$ . For instance, age is dichotomized and takes on values in  $\{0, 1\}$ , where 0 means that age  $< 80$ , and 1 means that age  $\geq 80$ . Each toxicity is categorized by integer-valued grades between 0 (absent) and either 3 or 4 (most severe). Let  $\mathbf{z}_{i,k}$  be the vector of all patient covariates in





**FIGURE 1** Example for one patient of dose adjustment according to the patient's characteristics and toxicities [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

$x_{i,k}$ , using the dummy coding of categorical and/or ordinal data, that is,  $z_{i,k}$  has length  $L = \sum_{j=1}^J C_j$  with each  $x_{i,k,j}$  split into  $C_j$  dummies variables, each taking a value of 1 if the corresponding category was selected in  $x_{i,k,j}$  and 0 otherwise. In the following section, we formally define clinical relevance weights and explain how we use them to determine prior distribution hyperparameters in a Bayesian model.

## 2.1 | Clinical relevance weights

Clinical relevant weights reflect the importance that a physician places on each value of each variable used in a medical decision. Once these weights are elicited, they can be used to compute the hyperparameters of the prior distributions of each variable's inclusion indicator in the Bayesian model. In the elicitation, each of several physicians who are experts in the field is asked to provide positive-valued numerical weights  $w_j$  for each variable  $j \in \{1, \dots, J\}$  and each of its possible values  $c \in \{1, \dots, C_j\}$ . Denote  $\mathbf{w}_j = (w_{j,1}, \dots, w_{j,C_j})$ , ordered such that  $0 = w_{j,0} \leq w_{j,1} \leq \dots \leq w_{j,C_j} = W_{\max}$  for each  $j$ . The value  $w_{j,c}$  reflects how much the expert thinks he/she would change the dose on a scale between 0 and  $W_{\max}$  when the severity level  $x_{j,c}$  is observed. The value  $w_{j,0}$  is not included in the weight vector because it represents the reference level, that is, no toxicities in the case of toxicity variables or the reference value in the case of a dichotomous variable. For example, letting  $W_{\max} = 100$ , the vector  $w_6 = \{0, 20, 80, 90\}$  refers to clinical relevance weights for the sixth variable, which has  $C_6 = 4$  levels. By analogy with  $\mathbf{z}$ , we define  $\mathbf{w}'$  to be the  $L$ -vector of all clinical relevance weights. This structure follows that structure introduced by Bekele and Thall,<sup>23</sup> who used elicited toxicity severity weights to define total toxicity burden for use in phase I dose-finding.

For our setting, oncologists helped us build a questionnaire by choosing the variables and thresholds that they used for decision making. Different experts could give different weights, which led to different prior distributions. Toxicity covariates were defined using the Common Terminology Criteria for Adverse Events' grades. Four clinicians separately completed the online questionnaire. Each clinician specified a numerical clinical relevance weight for each grade of each toxicity type and each level of each covariate, within the range  $[0, W_{\max}] = [0, 100]$ , with 0 corresponding to no severity and 100 the highest possible severity.

Table 1 presents the elicited clinical relevance weights for each oncologist and each variable. For example,  $\{0, 20, 80, 90\}$  in the row "vomiting" of Table 1 refers to clinical relevance weights of grades  $\{1, 2, 3, 4\}$  given by clinician 1. The table suggests that, if grade 2 vomiting is observed clinician 1 reduces the dose in 20% of cases. For all clinicians except clinician 1, the WHO score is very often considered for dose adjustment, and grade 4 is always considered for dose reduction. Clinician 1 gives the small weight of 20 to WHO scores 2, 3, and 4. Clinicians 2, 3, and 4 do not consider treatment line for decision making, while clinician 1 considers it for dose reduction. Concerning toxicities, vomiting is used very differently by the four clinicians to make decisions. Finally, on average, clinicians 2 and 3 appear to assign lower weights to many levels of many variables (cf sum of weights in Table 1) compared with clinicians 1 and 4. The elicited weights suggest that the four physicians are likely to act differently when facing the same situations.

## 2.2 | Weight-based SSVS

The general framework of clinical practice reflected by the Irinotecan data is repeated dose adaptation over successive treatment cycles based on the patient's most recently updated treatment, dose, and outcome data. In the first cycle, a starting dose is chosen based on the patient's baseline characteristics, including age, bilirubin, WHO score, and possibly previous treatments. Thereafter, if toxicities are observed in a given cycle, the clinician may adapt the dose for the following cycle. We assume the following linear mixed effect model, which is intended to use the updated covariates in each cycle

**TABLE 1** Clinical relevance weights for each variable elicited from each clinician

Covariates	Clinicians															
	1				2				3				4			
Age $\geq 80$ years	100	-	-	-	60	-	-	-	100	-	-	-	80	-	-	-
Treatment line 3, > 3	30	50	-	-	0	0	-	-	0	0	-	-	0	0	-	-
Weight loss > 10%	50	-	-	-	20	-	-	-	50	-	-	-	80	-	-	-
WHO score (1, 2, 3, 4)	0	20	20	20	0	0	40	100	0	0	80	100	0	20	80	100
Bilirubin >35 $\mu\text{mol/L}$	100	-	-	-	40	-	-	-	100	-	-	-	20	-	-	-
<b>Toxicity grades 1, 2, 3, 4</b>																
Vomiting	0	20	80	90	0	30	70	100	0	10	10	10	10	20	80	100
Nausea	0	20	80	-	0	10	50	-	0	10	10	-	10	30	80	-
Diarrhea	0	40	80	100	0	20	50	100	0	50	80	100	0	20	70	90
Asthenia	10	50	100	-	10	10	40	-	0	0	70	-	10	50	70	-
Neutropenia	0	70	100	100	0	0	30	50	0	0	50	50	0	20	70	80
Thrombopenia	40	100	100	100	0	0	20	30	0	0	50	50	0	50	80	100
Anemia	0	50	80	100	0	0	20	30	0	0	0	0	0	20	50	70
<b>Sum of weights</b>	1900				930				980				1560			

to explain the physician's decisions. Let  $d_{i,k}$  denote the dose given to patient  $i$  at cycle  $k$  and  $\mathbf{z}_{i,k}$  the vector of the patient's covariates at that cycle visit for  $i \in \{1, \dots, n\}$  and  $k \in \{1, \dots, K\}$ . The model is given by the equation

$$d_{i,k} = \theta_0 + \boldsymbol{\theta}^T \mathbf{z}_{i,k} + \gamma_i + \epsilon_{i,k}, \quad (1)$$

where  $\boldsymbol{\theta}$  is a parameter vector of length  $L$ ;  $\gamma_1, \dots, \gamma_n$  are iid  $\mathcal{N}(0, \sigma_\gamma^2)$  random patient effects; and  $\epsilon_{i,k} \sim \mathcal{N}(0, \sigma_\epsilon^2)$  is the error term. In this formulation, the Markov property is assumed conditional on the random effects such that the dose given at cycle  $k$  depends only on the covariates resulting from the previous cycle and collected in  $\mathbf{z}_{i,k}$ . Thus,  $\mathbf{z}_{i,k}$  includes indicators of the grades of any toxicities that occurred in cycle  $k - 1$  and any updated values of the baseline covariates. Note that toxicities are not the outcomes of the regression, but rather they are the covariates.

Our objective is to find a subset,  $\mathcal{I}$ , of the vector  $\mathbf{z}$  that describes the data well. For covariate index  $l \in \{1, \dots, L\}$ , let  $I_l$  to be the indicator variable that takes the values

$$I_l = \begin{cases} 1, & \text{if } z_{\cdot,l} \text{ is in the model} \\ 0, & \text{otherwise,} \end{cases}$$

where we refer to the  $l$ th covariate by  $z_{\cdot,l}$  for convenience. The SSVS approach assumes that the prior distribution for each covariate parameter in model (1) is a mixture of two normal distributions, both centered at zero but with different variances. Specifically, as given in the work of George and McCulloch,<sup>11</sup>

$$\theta_l | I_l \sim (1 - I_l) \mathcal{N}(0, \tau_l^2) + I_l \mathcal{N}(0, g_l \tau_l^2), \quad (2)$$

where  $\mathbb{P}(I_l = 1) = p_l$ , denoted by  $I_l \sim \mathcal{B}(p_l)$ . In Equation (2), the first term (the spike) accounts for the nonselected covariate  $z_{\cdot,l}$  ( $I_l = 0$ ), with the density centered at 0 having small variance. Note that the sequence of zero-centered normal distributions converges to the Dirac delta function at 0 as the variance converges to 0. Thus, rather than using a prior with a point mass on the regression coefficient equaling 0, we decided to use a normal distribution with small variance to approximate it and thus allow us to use usual statistical software. The second term accounts for the selected covariates  $z_{\cdot,l}$ , where  $I_l = 1$ , and therefore, it has a large variance.

In the classical SSVS framework,  $p_l$  is set to 0.5. Here, we use the elicited weights to choose a prior distribution for  $I_l$  such way that  $I_l$  is more likely to be 1 when its elicited clinical relevance weight is high, and it is more likely to be 0 otherwise. Rather than fixing  $p_l$ , we assume that it follows a beta prior distribution

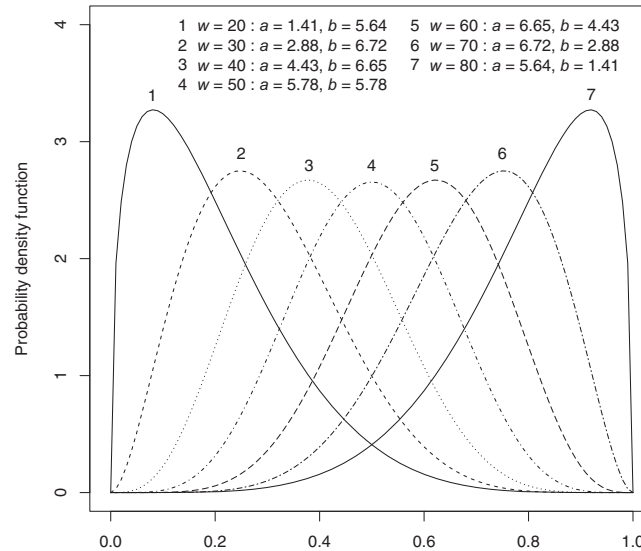
$$p_l \sim \text{Beta}(a_l, b_l). \quad (3)$$

Consequently,  $I_l | p_l$  follows a beta-binomial distribution with parameters  $(a_l, b_l)$ . The main innovation of our proposed WBS method is that we assume the prior equation given by Equation (3) and compute its hyperparameters  $a_l$  and  $b_l$  of using the elicited clinical relevance weights  $w'_l$  described in Section 2.1. Thus, first, each clinician separately specifies a



**TABLE 2**  $(a, b)$  values giving a beta( $a, b$ ) distribution centered at clinical relevance weight/100

Clinical Relevance Weights	0	10	20	30	40	50	60	70	80	90	100
$a$	1.41	1.41	1.41	2.88	4.43	5.78	6.65	6.72	5.64	5.64	5.64
$b$	5.64	5.64	5.64	6.72	6.65	5.78	4.43	2.88	1.41	1.41	1.41



**FIGURE 2** Probability density functions for  $p$ 's beta distributions with parameters  $a$  and  $b$  and mean  $w/100$

vector of numerical clinical relevance weights  $\mathbf{w}'$  associated with each grade of each toxicity type and each level of each covariate. Then, we find  $a_l$  and  $b_l$  by solving the equations

$$\begin{cases} \mathbb{E}(p_l) = \frac{a_l}{a_l+b_l} = \frac{w'_l}{W_{\max}} \\ \mathbb{V}(p_l) = \frac{a_l b_l}{(a_l+b_l)^2(a_l+b_l+1)} = S. \end{cases}$$

Thus, the mean of  $p_l$  is  $w'_l/W_{\max}$ , while  $S$  is a tuning parameter that accounts for uncertainty, chosen based on a preliminary sensitivity analysis. In our application, we choose the same  $S$  for all prior distributions, although in general, the value may be different for different  $p_l$ , reflecting different levels of the physician's confidence for different covariates. Note that no inference procedures are needed for the clinical relevance weights to be elicited from each clinician. To implement the WBS method, we set the minimal and maximal thresholds to 20 and 80, respectively, to avoid highly informative priors. Values of  $a$  and  $b$  obtained from clinical relevance weights used in the WBS method are presented in Table 2 and Figure 2. We choose  $W_{\max} = 100$  because the scale  $[0, 100]$  is well known and easily understandable by clinicians, and we set  $S = 0.02$  based on our preliminary sensitivity analysis.  $S$  quantifies the variability that one wants to add to the prior distribution. We denote the WBS method based on the  $m$ th clinician's clinical relevance weights as  $\text{WBS}(m)$ , for  $m \in \{1, 2, 3, 4\}$ .

### 2.3 | Combination of prior opinions

To combine prior opinions from different physicians, we propose two methods: (1) use of the WBS method with a mixture prior obtained by weighting each of the four elicited physician priors 0.25 each and (2) Bayesian Model Averaging (BMA) of the WBS models used with each clinician's clinical relevance weights. To accommodate inconsistent weighting by different physicians, BMA calculates the predictive distribution of a coefficient  $\theta_l, \in \{1, \dots, L\}$ , as a weighted average of posterior distributions of  $\theta_l$  using each model  $\text{WBS}(m), m \in \{1, \dots, 4\}$  (see the work of Raftery et al<sup>24</sup>)

$$p(\theta_l | \mathbf{d}, \mathbf{z}) = \sum_{m=1}^4 p(\text{WBS}(m) | \mathbf{d}, \mathbf{z}) p(\theta_l | \text{WBS}(m), \mathbf{d}, \mathbf{z}),$$

where the weights  $p(\text{WBS}(m)|\mathbf{d}, \mathbf{z})$  are the posterior probabilities of each model

$$p(\text{WBS}(m)|\mathbf{d}, \mathbf{z}) = \frac{p(\mathbf{d}, \mathbf{z}|\text{WBS}(m))p(\text{WBS}(m))}{\sum_{m_1=1}^4 p(\mathbf{d}, \mathbf{z}|\text{WBS}(m_1))p(\text{WBS}(m_1))}.$$

The marginal likelihood of the model  $\text{WBS}(m)$  is

$$p(\mathbf{d}, \mathbf{z}|\text{WBS}(m)) = \int p(\mathbf{d}, \mathbf{z}|\boldsymbol{\theta}, \text{WBS}(m))p(\boldsymbol{\theta}|\text{WBS}(m))d\boldsymbol{\theta},$$

where  $p(\mathbf{d}, \mathbf{z}|\boldsymbol{\theta}, \text{WBS}(m))$  is the likelihood that we can assess using the Laplace method and  $p(\boldsymbol{\theta}|\text{WBS}(m))$  the prior distribution. Then, the expected value and the variance of  $\theta_l$  are obtained by

$$\begin{cases} \mathbb{E}(\theta_l|\mathbf{d}, \mathbf{z}) = \sum_{m=1}^4 \hat{\theta}_l^m p(\text{WBS}(m)|D) \\ \mathbb{V}(\theta_l|\mathbf{d}, \mathbf{z}) = \sum_{m=1}^4 \left[ \mathbb{V}(\theta_l|\mathbf{d}, \mathbf{z}, \text{WBS}(m)) + \left( \hat{\theta}_l^m \right)^2 \right] p(\text{WBS}(m)|\mathbf{d}, \mathbf{z}) - \mathbb{E}(\theta_l|\mathbf{d}, \mathbf{z})^2, \end{cases}$$

where  $\hat{\theta}_l^m = \mathbb{E}(\theta_l|\mathbf{d}, \mathbf{z}, \text{WBS}(m))$ .

## 2.4 | Model comparison

In our simulations, we compare the performance of the WBS method to the classical Bayesian SSVS and the frequentist Lasso. Our evaluation criteria are bias, pointwise prediction performance, the posterior distribution of the prediction performance, and variable selection performance.

### Bias

The bias of the estimator  $\hat{\theta}_l, l \in \{1, \dots, L\}$ , is defined as the difference  $\text{Bias}(\hat{\theta}_l) = \mathbb{E}(\hat{\theta}_l) - \theta_l$  between the expected value of the estimator and the true values of the coefficient  $\theta_l$ , where  $\hat{\theta}_l$  is the posterior mean of  $\theta_l$  for the Bayesian methods and the Lasso estimates otherwise.

### Performance of prediction

To evaluate the prediction performance of each method, we use the root mean square error (RMSE), computed as the sample standard deviation of the differences between predicted and observed values

$$\text{RMSE}(\hat{\mathbf{d}}) = \sqrt{\text{MSE}(\hat{\mathbf{d}})} = \sqrt{\mathbb{E}((\hat{\mathbf{d}} - \tilde{\mathbf{d}})^2)} \approx \sqrt{\frac{1}{n} \frac{1}{K} \sum_{i=1}^n \sum_{k=1}^K (\hat{d}_{i,k} - \tilde{d}_{i,k})^2},$$

where  $\tilde{\mathbf{d}}$  are observed doses for patients in the validation data set and  $\hat{\mathbf{d}} = \hat{\boldsymbol{\theta}}_0 + \hat{\boldsymbol{\theta}}^T \tilde{\mathbf{Z}}$  are the estimations given by the model for these doses. Additional details are presented in Appendix A.

### Bayesian predictive model accuracy

Unlike frequentist methods, the Bayesian approach provides a posterior distribution, which provides a basis for computing the log pointwise predictive density, proposed by Gelman et al.<sup>25</sup> We adapt this approach to the repeated measures structure of the Irinotecan data set. Denoting the prior distribution as  $p(\boldsymbol{\theta}, \mathbf{I}, \boldsymbol{\gamma}, \sigma_\gamma^2, \sigma_\epsilon^2)$  and the posterior distribution as  $p_{\text{post}}(\boldsymbol{\theta}, \mathbf{I}, \boldsymbol{\gamma}, \sigma_\gamma^2, \sigma_\epsilon^2 | \text{data})$ , the posterior predictive distribution of  $\tilde{\mathbf{d}}$  is

$$p_{\text{post}}(\tilde{\mathbf{d}} | \text{data}) = \int p(\tilde{\mathbf{d}}|\boldsymbol{\theta}, \mathbf{I}, \boldsymbol{\gamma}, \sigma_\gamma^2, \sigma_\epsilon^2) p_{\text{post}}(\boldsymbol{\theta}, \mathbf{I}, \boldsymbol{\gamma}, \sigma_\gamma^2, \sigma_\epsilon^2 | \text{data}) d\boldsymbol{\theta} d\mathbf{I} d\boldsymbol{\gamma} d\sigma_\gamma^2 d\sigma_\epsilon^2.$$

In practice, the distribution  $f_{i,k}(\mathbf{d})$  depends on  $(\boldsymbol{\theta}, \mathbf{I}, \boldsymbol{\gamma}, \sigma_\gamma^2, \sigma_\epsilon^2)$ , which are not known. We thus work with the log pointwise predictive density

$$\text{LPD} = \sum_{i=1}^n \sum_{k=1}^K \log p_{\text{post}}(\tilde{d}_{i,k}) = \sum_{i=1}^n \sum_{k=1}^K \log \int p(\tilde{d}_{i,k}|\boldsymbol{\theta}, \mathbf{I}, \boldsymbol{\gamma}, \sigma_\gamma^2, \sigma_\epsilon^2) p_{\text{post}}(\boldsymbol{\theta} | \text{data}) d\boldsymbol{\theta} d\mathbf{I} d\boldsymbol{\gamma} d\sigma_\gamma^2 d\sigma_\epsilon^2.$$

To assess the predictive accuracy for the  $n \times K$  data points in a data set, the expected log pointwise predictive density (ELPD) is defined as follows:

$$\text{ELPD} = \sum_{i=1}^n \sum_{k=1}^K E_{f_{i,k}} \{\log p_{\text{post}}(\tilde{d}_{i,k})\},$$

where  $f_{i,k}(\mathbf{d})$  is the distribution representing the true data-generating process for  $d_{i,k}$ .

For computational details, see Appendix B. Because each model has the same number of parameters, and they are estimated in the same way, one might simply compare their best-fit log predictive densities directly. We thus select the model with the smallest value of  $-\text{LPD}$ .

### Precision of variable selection

A method's variable selection ability may be assessed through its false positive rate (FPR) and false negative rate (FNR). The FPR is defined to be the number of nonpredictive variables, ie, having true coefficient 0, that are wrongly selected by the method (false positive = FP) divided by the total number of actual nonpredictive variables (N)

$$\text{FPR} = \frac{\text{FP}}{\text{N}} = \frac{\text{FP}}{\text{FP} + \text{TN}},$$

where  $\text{TN}$  is the number of true negatives, ie, covariates having true coefficient 0. In the Bayesian framework, a variable is considered selected when its posterior median is equal to one. For Lasso methods, all nonselected covariates are estimated as exactly zero. Note that  $\text{FPR} = 1 - \text{TNR}$ , where  $\text{TNR}$  is the true negative rate or specificity.

Similarly, the FNR is defined as

$$\text{FNR} = \frac{\text{FN}}{\text{P}} = \frac{\text{FN}}{\text{FN} + \text{TP}},$$

where  $\text{FN}$  is the number of false negatives,  $\text{TP}$  is the number of true positives, and  $\text{P} = \text{FN} + \text{TP}$  is the total number of positive selections. Additionally,  $\text{FNR} = 1 - \text{TPR}$  where  $\text{TPR}$  is the true positive rate or sensitivity.

## 3 | SIMULATION DESIGN

We simulated a sample of patients treated for colorectal metastatic cancer by Irinotecan, to imitate the structure of our motivating data set. The protocol is based on a theoretical dose of 180 mg/m<sup>2</sup>. Severity levels of toxicities were simulated with grades taking on integer values from 0 (indicating no toxicity of that type) to 4 (the most severe level), or to 3 for nausea and asthenia.

### 3.1 | Covariate distributions used in the simulations

Numerical parameter choices for the simulations were based on case study covariate distributions (cf Table 3).  $X_{.1}$  and  $X_{.2}$  are assumed to be fixed in time. All other variables change over time and were simulated with an autocorrelation structure of order 1 ( $\rho$ ) between cycles according to a Gaussian copula-based procedure.

### 3.2 | Dose generation

The dose for individual  $i \in \{1, \dots, n\}$  at cycle  $k \in \{1, \dots, K\}$  was generated from the model

$$d_{i,k} = D_{\text{max}} - \boldsymbol{\beta}^T \mathbf{z}_{i,k} + \epsilon_{i,k},$$

where  $\epsilon_{i,k} \sim \mathcal{N}(0, \sigma^2)$ . Four realistic scenarios were simulated: two involving clinician 1's clinical relevance weights (scenarios 1a and 1b) and two involving clinician 2's clinical relevance weights (scenarios 2a and 2b). First, we selected variables with clinical relevance weights greater than specific thresholds; then, we set coefficients  $\boldsymbol{\beta}$  of the selected variables as a linear function of the clinical relevance weights such that all doses were included between  $D_{\text{min}}$  and  $D_{\text{max}}$ . We also constructed an unrealistic scenario where the clinical relevance weights used to simulate the scenario are taken equal to  $100 - \mathbf{w}'$  for clinician 1. Thus, for example, vomiting of grade 4 is linked to a smaller clinical weight that vomiting of grade 3. So, considering priors built on clinician expertise could favor inclusion of wrong variables. Values of the thresholds and coefficients  $-\boldsymbol{\beta}$  are given in Table 4.

**TABLE 3** Simulation parameters for the covariates

Variables	Original Variables	Probability Distribution	Parameters
$x_{.1}$	Age $\geq 80$ years	$\mathcal{N}(\mu, s^2)$	$\mu = 63, s = 12$
$x_{.2}$	Treatment line 3, $> 3$	$Geom(r_1)$	$r_1 = 0.5$
$x_{.3}$	Weight loss $> 10\%$	$B(r_2)$	$r_2 = 0.1$
$x_{.4}$	WHO score	$B(l, r_3)$	$l = 4, r_3 = 0.5$
$x_{.5}$	Bilirubin $> 35 \mu\text{mol/L}$	$B(r_3)$	$r_4 = 0.1$
$x_{.6}$	Vomiting	$Geom(r_5)$ truncated to $s_1$	$r_5 = 0.2, s_1 = 4$
$x_{.7}$	Nausea	$Geom(r_6)$ truncated to $s_2$	$r_6 = 0.2, s_2 = 3$
$x_{.8}$	Diarrhea	$Geom(r_7)$ truncated to $s_3$	$r_7 = 0.2, s_3 = 4$
$x_{.9}$	Asthenia	$Geom(r_8)$ truncated to $s_4$	$r_8 = 0.2, s_4 = 3$
$x_{\dots,10}$	Neutropenia	$Geom(r_9)$ truncated to $s_5$	$r_9 = 0.2, s_5 = 4$
$x_{\dots,11}$	Thrombopenia	$Geom(r_{10})$ truncated to $s_6$	$r_{10} = 0.2, s_6 = 4$
$x_{\dots,12}$	Anemia	$Geom(r_{11})$ truncated to $s_7$	$r_{11} = 0.2, s_7 = 4$

Abbreviation: WHO, World Health Organization.

**TABLE 4** Values of  $-\beta$  used for each simulation scenario

Variable	Realistic Scenarios				Unrealistic Scenario
	1a	1b	2a	2b	F
$X_{.1}$	18.6	13.4	14.4	12.6	0
$X_{.2}$	0, 0	0, 6.70	0, 0	0, 0	0, 0
$X_{.3}$	0	6.70	0	4.19	0
$X_{.4}$	0, 0, 0, 0	0, 0, 0, 0	0, 0, 9.63, 24.1	0, 0, 8.39, 21.0	18.8, 15.1, 15.1, 15.1
$X_{.5}$	18.6	13.4	9.63	8.39	0
$X_{.6}$	0, 0, 0, 0	0, 0, 10.7, 12.1	0, 0, 16.9, 24.1	0, 6.29, 14.7, 21.0	18.8, 15.1, 0, 0
$X_{.7}$	0, 0, 0	0, 0, 10.7	0, 0, 12.0	0, 0, 10.5	18.8, 15.1, 0
$X_{.8}$	0, 0, 0, 18.6	0, 0, 10.7, 13.4	0, 0, 12.0, 24.1	0, 4.19, 10.5, 21.0	18.8, 0, 0, 0
$X_{.9}$	0, 0, 18.6	0, 6.70, 13.4	0, 0, 9.63	0, 0, 8.39	17.0, 0, 0
$X_{\dots,10}$	0, 0, 18.6, 18.6	0, 9.39, 13.4, 13.4	0, 0, 0, 12.0	0, 0, 6.29, 10.5	18.8, 0, 0, 0
$X_{\dots,11}$	0, 18.6, 18.6, 18.6	0, 13.4, 13.4, 13.4	0, 0, 0, 0	0, 0, 4.19, 6.29	0, 0, 0, 0
$X_{\dots,12}$	0, 0, 0, 18.6	0, 6.70, 10.7, 13.4	0, 0, 0, 0	0, 0, 4.19, 6.29	18.8, 0, 0, 0
Number of variables actually used	10	20	11	19	12
Weights' threshold	100	50	40	20	80

To assess the extent to which the prior deviates from the simulated parameter value, we calculate Spearman's rank correlation coefficients between the following:

- corrected clinical relevance weights used to build prior distributions ( $w^c$ ), that is, clinical relevance weights with minimal and maximal thresholds to 20 and 80 respectively;
- weights used to build scenarios ( $w^u$ ), that is, only clinical relevance weights greater than thresholds specific to each scenario are nonzero weights.

It allows one to measure the extent to which, as one variable increases, the other variable tends to increase. Table 5 gives Spearman coefficient between each model and each scenario.

For each of the five scenarios, we studied three cases: ( $n = 10, K = 3$ ), ( $n = 10, K = 5$ ), and ( $n = 20, K = 5$ ). Values of parameters were derived from case study covariates (cf Table 6). We set  $\sigma = 5$  based on suggestions reported in the literature.

For each of the 15 subscenarios, 1000 data sets were simulated. Each data set consisted of a training set of size  $n \times K$ , along with an independent validation set of size  $50 \times 5$ . For each simulated data set, we fit eight models on the training data sets using the Lasso method, the classical SSVS method, and the WBS method with clinical relevance weights provided by each of the four clinicians, with mixture and with BMA. All analyses were performed using R software 3.3.2 version

**TABLE 5** Spearman's rank correlation coefficient between corrected clinical relevance weights (in lines) and weights used to build scenarios (in columns)

WBS	Scenarios				F
	1a	1b	2a	2b	
1	0.668	0.948	0.426	0.585	-0.857
2	0.379	0.505	0.913	0.949	-0.318
3	0.520	0.424	0.562	0.669	-0.300
4	0.436	0.610	0.655	0.788	-0.454

Abbreviations: WBS, weight-based SSVS.

**TABLE 6** Parameter values in the simulations

Parameter	Values
$n$	10 or 20
$K$	3 or 5
$J$	12
$L$	35
$\rho$	0.5
$D_{\min}$	50
$D_{\max}$	180
$\sigma$	5
S	0.02
$W_{\max}$	100

and packages *glmmLasso* and *R2jags*. In *R2jags*, *autojags* was implemented with 3 chains and a burn-in of 1000 and, as a convergence criterion, Gelman and Rubin's potential scale reduction factor  $Rhat = 1.1$ .

### 3.3 | Prior parameter settings

In Bayesian estimation of mixed models (cf Section 2.2, Equation (1)), prior distributions for the parameters  $\theta$ ,  $\sigma_\gamma$ , and  $\sigma_\epsilon$  must be defined. After a sensitivity analysis in which several parameter values were tested and model performance was evaluated in a few scenarios, an inverse-gamma distribution with shape and scale parameters (1, 1) and (0.2, 1), respectively, were chosen for the covariance of the random effects  $\sigma_\gamma^2$  and the residual variance  $\sigma_\epsilon^2$ , (cf Table 7). For  $\theta$ , see Section 2.2.

The models used by the SSVS and WBS algorithms also require specification of fixed prior hyperparameters,  $\tau^2$  and  $\mathbf{g}\tau^2$ , which we based on the data (cf Equation (2)). For  $l \in \{1, \dots, L\}$ ,  $\tau_l$  should be chosen such that, if  $\theta_l \sim \mathcal{N}(0, \tau_l^2)$ , then  $\theta_l$  can be "safely" replaced by 0. Moreover,  $g_l (>1)$  should be chosen such that if  $\theta_l \sim \mathcal{N}(0, g_l \tau_l^2)$ , then a nonzero estimate of  $\theta_l$  should be included in the final model. Some suggestions can be found in the work of George and McCulloch.<sup>11</sup> Because the smallest numerical coefficient value for our simulated data is 4.19 (cf Table 4), we take  $3\tau_l$  to be equal to the maximum value at which  $\theta_l$  would be equivalent to 0. In the following, based on our sensitivity analysis, we assume  $\tau_l = 1$  and  $g_l = 100^2$ .

**TABLE 7** Choice of prior distributions

Parameter	Prior Distribution	Hyperparameters
$\sigma_\gamma^2$	Inverse-Gamma( $\alpha_1, \beta_1$ )	$\alpha_1 = 1, \beta_1 = 1$
$\sigma_\epsilon^2$	Inverse-Gamma( $\alpha_2, \beta_2$ )	$\alpha_2 = 0.2, \beta_2 = 1$
$\theta_0$	$N(\mu_0, s_0^2)$ truncated to 0	$\mu_0 = 100, s_0 = 100$

## 4 | SIMULATION RESULTS

### 4.1 | Bias

The coefficients' bias (cf Figures 1 and 2 in supplementary material) are on average negative and dramatically decrease with their real values; therefore, the methods underestimate the coefficients of variables actually used. For realistic scenarios, the coefficients of variables actually not used are estimated to be no more than  $-3 \text{ mg/m}^2$ , that is,  $-1.67\%$  of the baseline dose.

### 4.2 | Prediction performance

Tables 8 and 9 summarize the average prediction performance (RMSEs) for validation sets over 1000 runs, along with the precision of selected variables (mean FPR and mean FNR) for each method. Figure 3 shows the distribution of RMSE for all scenarios with  $n = 10$  and  $K = 3$ .

Recall that the weights of clinicians 1 and 2 were used for scenarios 1a and 1b and scenarios 2a and 2b, respectively, in the simulations. For scenarios 1a and 1b, WBS1 has a slightly smaller RMSE than that of SSVS and the WBS2, WBS3, and WBS4 when  $n = 10$  patients and  $K = 3$  cycles (cf Table 8 and Figures 3A and 3B). However, the improvement is so small that only a mean improvement of less than  $2.3 \text{ mg/m}^2$  in the RMSE (for a baseline dose of  $180 \text{ mg/m}^2$ ) compared with the SSVS method is observed. The performance of WBS4 is similar to that of WBS1 as the Spearman's rank correlation coefficient between scenarios 1a/1b and  $\text{WBS}(m)$  takes the highest values for WBS1 and WBS4 (cf Table 5). WBS with the mixture does not improve prediction performance, compared to WBS with BMA that manages inconsistent weighting by the different physicians by favoring WBS1 and WBS4 models.

For scenarios 2a and 2b, the performance is approximately the same for all methods considered (cf Table 8 and Figures 3C and 3D). The performance of WBS is similar to that of SSVS, regardless of the weights used for the prior. For  $n = 10$  patients with  $K = 3$  cycles, the best performance is not achieved with WBS2, as might have been expected, but rather is best with the WBS4 and WBS with BMA. The WBS4's RMSE decreases by approximately  $1.7 \text{ mg/m}^2$  (for a baseline dose of  $180 \text{ mg/m}^2$ ), when  $n = 10$  patients and  $K = 3$  cycles, compared with that of the SSVS method.

Scenario F is the unrealistic scenario built using the importance weights  $\{100 - \text{clinical relevance weights of clinician 1}\}$ . For this scenario, all WBS methods perform poorly and the smallest RMSE is obtained for SSVS model (cf Table 9 and Figure 3E). Although all Spearman coefficients are negative between scenario F and  $\text{WBS}(m)$ ,  $m \in \{1, 2, 3, 4\}$ , the higher Spearman coefficients for this scenario are obtained for WBS3 and WBS2 models for which we find the best performances among the WBS models (cf Table 5).

Table 1 and Figure 3 in supplementary material present the  $-1 \times$  mean log pointwise predictive density over 1000 replications for each method. The log pointwise predictive density between the different methods is similar, as is the case for RMSE.

In general, the frequentist LASSO shows worse performance except for the unrealistic scenario. Performance improves with an increase in sample size and increased number of cycles. The WBS method predicts no better than the other methods for  $n = 20$ . The RMSE is close to the variance of the error term ( $\sigma = 5$ ) for all methods.

### 4.3 | Precision of selected variables

Figure 4 and Figures 4 and 5 in the supplementary material respectively show the percentage of times in which each variable is selected by the model, the true positive rate plotted against the FPR and the percentage of times versus real coefficients' values when  $n = 10$  and  $K = 3$ . For realistic scenarios, the WBS method performs better than the other methods in terms of variable selection, with both lower FPRs and FNRs in general. The SSVS and WBS methods select a few variables not actually used (less than 5% FPR), but they fail more often to correctly select variables actually used, with up to 84% FNR. For small patient samples, the Lasso shows very large FNRs. For scenarios 1a, 1b, 2a, and 2b, the FNRs of the Lasso are 65%, 81%, 67%, and 79%, respectively. When the number of observations increases, the Lasso wrongly selects many variables, with FPRs of 45%, 85%, 56%, and 80%, while the SSVS and WBS methods only select the true variables with FPR and FNR values close to 0.

Again, for scenarios 1a and 1b, the WBS1 model gives the best performance in selecting variables, followed by WBS with BMA and WBS4 (cf Table 8, Figures 4A and 4B, and Figures 4a, 4b, 5a, and 5b in the supplementary material). Indeed, WBS1 improves the selection ability, with 46% and 30% smaller FNR in scenarios 1a and 1b, respectively, compared with

**TABLE 8** Simulation results for realistic scenarios: mean root mean square errors (RMSEs) and mean false positive rate (FPR) and mean false negative rate (FNR) over 1000 replications

	Lasso	Classical SSVS	WBS					
			1	2	3	4	Mixture	BMA
<b>Scenario 1a: (n,K) = (10,3)</b>								
RMSE	13.4 (1.58)	11.3 (1.83)	8.99 (1.77)	11.8 (1.57)	10.9 (1.65)	10.0 (1.94)	10.4 (1.85)	9.11 (1.87)
FPR	12	3	3	3	3	4	2	3
FNR	65	62	16	68	53	33	45	17
<b>Scenario 1a: (n,K) = (10,5)</b>								
RMSE	9.46 (2.00)	7.32 (1.46)	6.53 (0.907)	8.07 (1.87)	7.15 (1.38)	6.78 (1.05)	6.84 (1.17)	6.59 (0.920)
FPR	37	1	1	1	1	1	1	1
FNR	23	13	1	22	9	4	5	1
<b>Scenario 1a: (n,K) = (20,5)</b>								
RMSE	6.17 (0.674)	5.50 (0.558)	5.63 (0.563)	5.63 (0.574)	5.63 (0.577)	5.64 (0.578)	5.62 (0.572)	5.63 (0.565)
FPR	45	0	0	0	0	0	0	0
FNR	2	1	0	0	0	0	0	0
<b>Scenario 1b: (n,K) = (10,3)</b>								
RMSE	16.5 (1.46)	14.6 (1.82)	12.9 (1.94)	14.6 (1.55)	14.6 (1.62)	13.8 (1.93)	14.0 (1.74)	13.1 (2.11)
FPR	11	5	1	4	4	4	3	2
FNR	81	84	54	84	81	66	76	55
<b>Scenario 1b: (n,K) = (10,5)</b>								
RMSE	13.1 (2.14)	12.9 (1.41)	9.79 (1.54)	13.3 (1.23)	13.2 (1.25)	11.4 (1.54)	12.2 (1.44)	9.81 (1.56)
FPR	42	3	0	2	3	2	2	0
FNR	41	73	33	76	73	50	63	33
<b>Scenario 1b: (n,K) = (20,5)</b>								
RMSE	6.92 (0.774)	6.33 (0.708)	6.20 (0.624)	6.78 (1.10)	6.75 (1.03)	6.34 (0.688)	6.37 (0.693)	6.22 (0.626)
FPR	85	0	0	0	0	0	0	0
FNR	3	9	7	15	14	9	10	7
<b>Scenario 2a: (n,K) = (10,3)</b>								
RMSE	14.1 (1.84)	11.1 (2.00)	10.2 (1.83)	9.79 (1.54)	10.6 (1.92)	9.34 (1.72)	9.72 (1.74)	9.54 (1.83)
FPR	12	3	4	0	1	2	1	3
FNR	67	58	36	44	44	29	38	29
<b>Scenario 2a: (n,K) = (10,5)</b>								
RMSE	9.50 (1.74)	7.62 (1.27)	7.22 (1.11)	7.38 (1.12)	7.19 (1.20)	7.03 (1.00)	7.12 (1.08)	6.97 (1.01)
FPR	39	1	1	0	0	1	0	1
FNR	24	23	11	19	12	11	13	8
<b>Scenario 2a: (n,K) = (20,5)</b>								
RMSE	6.37 (0.704)	5.72 (0.566)	5.70 (0.559)	5.71 (0.556)	5.67 (0.559)	5.72 (0.559)	5.68 (0.553)	5.68 (0.544)
FPR	56	0	0	0	0	0	0	0
FNR	3	2	1	2	1	2	1	0
<b>Scenario 2b: (n,K) = (10,3)</b>								
RMSE	14.9 (1.66)	12.6 (1.82)	11.8 (1.84)	11.5 (1.64)	12.3 (1.72)	10.9 (1.77)	11.5 (1.72)	11.3 (1.92)
FPR	12	4	2	1	1	1	1	2
FNR	79	79	61	68	72	58	68	58
<b>Scenario 2b: (n,K) = (10,5)</b>								
RMSE	11.2 (1.97)	9.84 (1.42)	9.00 (1.25)	9.54 (1.19)	9.61 (1.43)	8.70 (1.15)	9.13 (1.25)	8.70 (1.17)
FPR	39	2	1	0	0	0	0	1
FNR	45	63	47	60	57	46	54	44
<b>Scenario 2b: (n,K) = (20,5)</b>								
RMSE	6.96 (0.730)	6.65 (0.664)	6.40 (0.621)	6.83 (0.659)	6.62 (0.638)	6.49 (0.624)	6.57 (0.634)	6.39 (0.605)
FPR	80	0	0	0	0	0	0	0
FNR	8	31	23	37	32	26	30	23

Standards deviations are shown in parentheses.

Abbreviations: BMA, Bayesian Model Averaging; SSVS, stochastic search variable selection; WBS, weight-based SSVS.



**TABLE 9** Simulation results for the unrealistic scenario: mean root mean square errors (RMSEs) and mean false positive rate (FPR) and mean false negative rate (FNR) over 1000 replications

	Lasso	Classical SSVS	WBS					
			1	2	3	4	Mixture	BMA
<b>Scenario F: (n,K) = (10,3)</b>								
RMSE	18.2 (3.26)	11.9 (3.78)	20.2 (3.53)	16.8 (3.46)	18.0 (3.61)	19.3 (3.69)	17.9 (3.53)	18.6 (4.53)
FPR	16	5	23	6	8	16	9	18
FNR	60	49	88	78	79	82	82	79
<b>Scenario F: (n,K) = (10,5)</b>								
RMSE	10.3 (2.27)	6.97 (0.872)	8.47 (2.60)	7.45 (1.69)	7.38 (1.84)	7.75 (2.18)	7.59 (1.78)	7.40 (1.72)
FPR	46	2	6	3	3	4	3	4
FNR	23	24	36	29	27	28	30	26
<b>Scenario F: (n,K) = (20,5)</b>								
RMSE	6.92 (0.678)	5.72 (0.567)	6.06 (0.663)	5.77 (0.614)	5.68 (0.553)	5.71 (0.555)	5.78 (0.598)	5.69 (0.546)
FPR	70	1	2	1	1	1	1	1
FNR	10	5	16	9	6	7	9	6

Standards deviations are shown in parentheses.

Abbreviations: BMA, Bayesian Model Averaging; SSVS, stochastic search variable selection; WBS, weight-based SSVS.

the SSVS method, when  $n = 10$  and  $K = 3$ . In the simulations, recall that clinical relevance weights were used (1) to choose what variables are selected and (2) to estimate the coefficients associated with the variables used for dose reduction. Therefore, the ideal clinical relevance weights used in the models should all be either 0 or 100, implying that clinicians are either sure to use a covariate or not in his/her decision process. Scenario 1a is an example in which all covariates that have an influence on doses are associated with a clinical relevance weight of 100 for the WBS1 method. For scenarios 2a and 2b, like WBS with BMA, the WBS4 and WBS1 models give low FPR and FNR because of their higher clinical relevance weights (cf Table 8, Figures 4C and 4D, and Figures 4c, 4d, 5c, and 5d in the supplementary material). In these two scenarios, the WBS4 model yields 29% and 58% FNR, respectively, when  $n = 10$  and  $K = 3$ , while the SSVS method yields much larger FNR values 58% and 79%. Overall, WBS1 and WBS4 are associated with larger clinical relevance weights than WBS2 and WBS3 are (cf sum of clinical relevance weights for each clinician in Table 1), and they select a higher number of covariates. Interestingly, the WBS model that selects the highest number of covariates actually used, and a smaller number of covariates actually not used, is not necessarily the one that uses the clinical relevance weights of the clinician on which the simulation setting is based.

Furthermore, WBS1 fails to select a higher percentage of variables for scenario 1b, with an FNR of 54% for  $(n, K) = (10, 3)$ , compared with scenario 1a, with an FNR of 16% for  $(n, K) = (10, 3)$ , generated based on fewer variables and larger coefficients for the variables' effect on selected doses. Similarly, scenario 2a has a lower FNR than scenario 2b. Finally, for  $n = 20$  patients, the FNR values are close to 0 for scenario 2a but are high for scenario 2b. As expected, the ability to correctly select variables increases with increasing sample size and number of cycles.

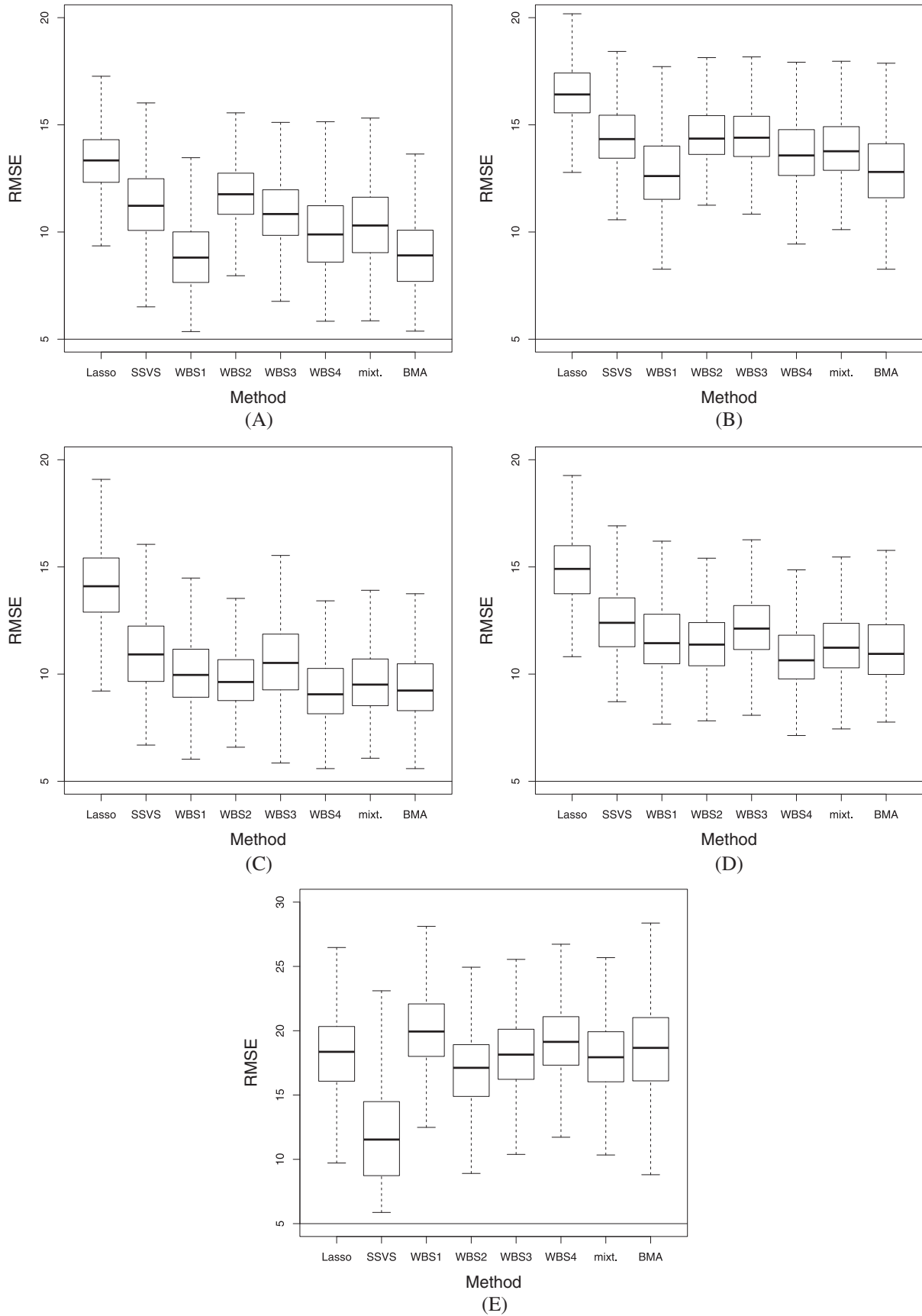
For scenario F, the variable selection performance of our method is as bad as the predictive performance for this scenario (cf Table 9, Figure 4E, and Figures 4e and 5e in the supplementary material).

The Lasso failed to converge in 1 to 11% of the simulated cases, depending on the scenario, with a median of 6%, while the other methods failed to converge in less than 2% of the cases.

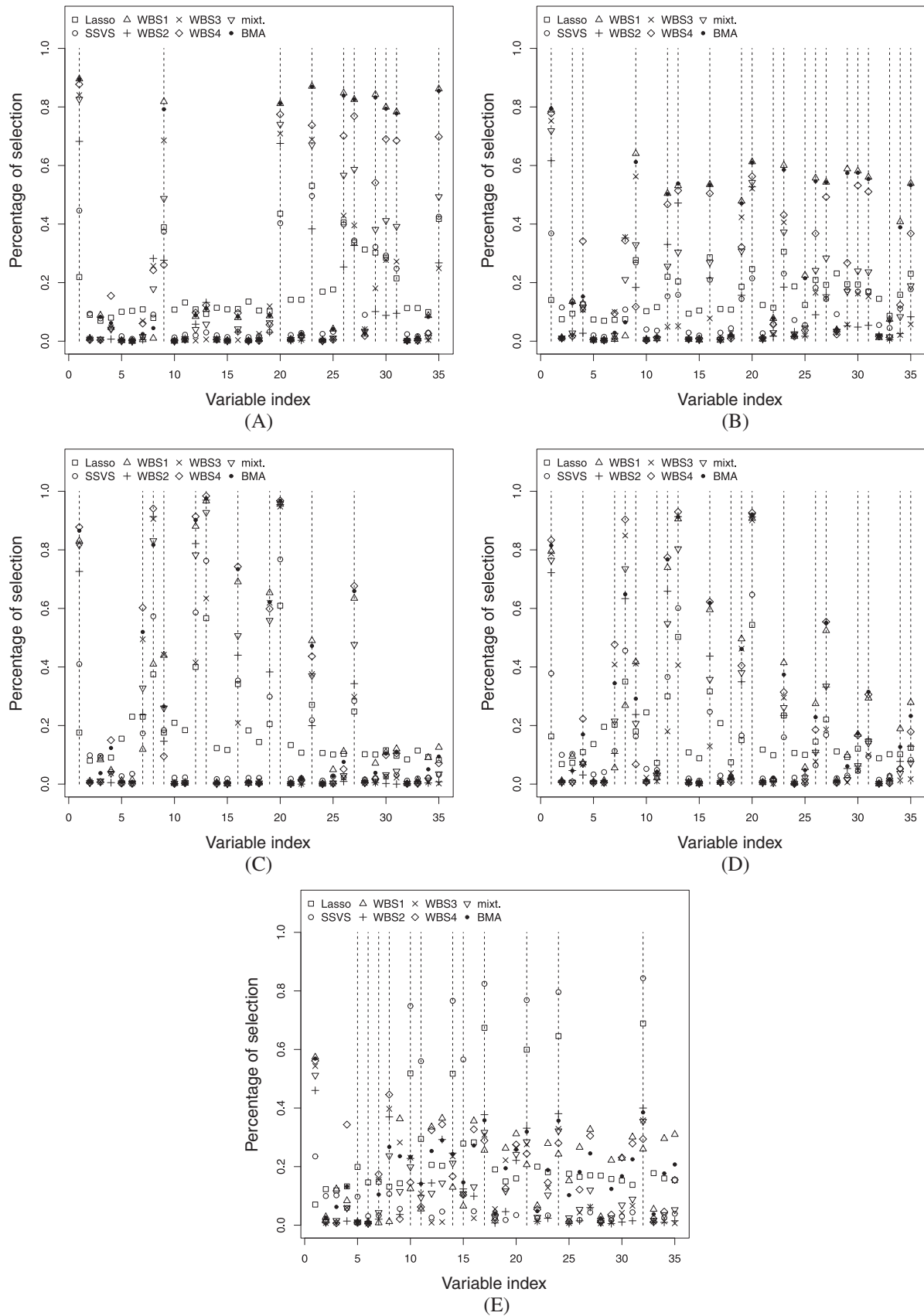
## 5 | CASE STUDY

We extracted data for patients treated for metastatic colorectal cancer with a combination of drugs including Irinotecan from the Georges Pompidou University Hospital (HEGP) I2B2 warehouse with IRB approval. The protocol encompasses a theoretical dose of 180 mg/m<sup>2</sup> and a theoretical cycle of chemotherapy of 14 days. All patients who were treated with any other protocol including Irinotecan were excluded. One cycle was defined as doses being given over one to three successive days. When a patient received a dose more than 28 days later, we considered that to be the beginning of another protocol. The data included the covariates age, weight, total bilirubin, WHO score, treatment line, and the seven toxicity types registered on each cycle of chemotherapy, as described in Section 3, using the physicians' elicited clinical relevance weights.





**FIGURE 3** Boxplot of root mean square error (RMSE) for each method and for each scenario with  $n = 10$  and  $K = 3$ . A, Scenario 1a:  $(n, K) = (10, 3)$ ; B, Scenario 1b:  $(n, K) = (10, 3)$ ; C, Scenario 2a:  $(n, K) = (10, 3)$ ; D, Scenario 2b:  $(n, K) = (10, 3)$ ; E, Scenario F:  $(n, K) = (10, 3)$



**FIGURE 4** Percentage of times where each variable is selected by the model for scenarios with  $n = 10$  and  $K = 3$ . Variables actually used are identified by dashed lines. A, Scenario 1a:  $(n, K) = (10, 3)$ ; B, Scenario 1b:  $(n, K) = (10, 3)$ ; C, Scenario 2a:  $(n, K) = (10, 3)$ ; D, Scenario 2b:  $(n, K) = (10, 3)$ ; E, Scenario F:  $(n, K) = (10, 3)$

**TABLE 10** Case study: description of covariates

Variables	0	1	2	3	4
Age ≥ 80 years	19 (58)	14 (42)	-	-	-
Weight loss > 10%	32 (97)	1 (3)	-	-	-
WHO score	7 (21)	21 (64)	5 (15)	-	-
Bilirubin >35 μmol/L	33 (100)	-	-	-	-
Treatment line 3, > 3	33 (100)	-	-	-	-
<b>Toxicity Grades</b>					
Vomiting	33 (100)	-	-	-	-
Nausea	15 (45)	16 (48)	2 (6)	-	-
Diarrhea	17 (52)	15 (45)	1 (3)	-	-
Asthenia	4 (12)	18 (55)	10 (30)	1 (3)	-
Neutropenia	28 (85)	3 (9)	2 (6)	-	-
Thrombopenia	28 (85)	5 (15)	-	-	-
Anemia	23 (70)	9 (27)	1 (3)	-	-

Percentages are shown in parentheses.

Abbreviation: WHO, World Health Organization.

**TABLE 11** Case study results: estimated coefficients, root mean square errors and -log pointwise predictive density

Variables	Weights of Clinicians:				Lasso	Classical SSVS	WBS With Prior From Clinicians:					BMA
	1	2	3	4			1	2	3	4	Mixture	
Intercept	-	-	-	-	178	174	176	173	178	174	175	176
Age ≥ 80 years	100	60	100	80	-11.1*	-3.21	-10.1*	-5.56	-13.4*	-8.79*	-10.2*	-11.0*
Weight loss > 10%	50	20	50	80	0	-0.267	-0.664	-0.177	-0.737	-1.72	-0.625	-0.850
WHO score 1	0	0	0	0	-10.9*	0.268	-0.00462	0.0618	0.194	0.0936	0.0663	0.0895
WHO score 2	20	0	0	20	0	-0.721	-0.454	-0.590	-0.754	-0.438	-0.522	-0.570
Nausea 1	0	0	0	10	0	-0.337	-0.342	-0.605	-2.16	-1.95	-0.925	-1.31
Nausea 2	20	10	10	30	0	0.483	0.305	2.32	2.24	1.89	1.35	1.38
Diarrhea 1	0	0	0	0	0	-0.784	-1.44	-0.399	-2.34	-0.268	-0.714	-1.55
Diarrhea 2	40	20	50	20	0	1.23	0.313	0.222	0.175	0.180	0.191	0.235
Asthenia 1	10	10	0	10	0	-0.238	0.0724	0.0915	0.0812	0.0534	0.0955	0.0733
Asthenia 2	50	10	0	50	7.39*	-6.57*	-2.54	-1.61	-3.36	-1.58	-1.88	-2.65
Asthenia 3	100	40	70	70	0	-86.7*	-82.1*	-80.5*	-85.3*	-81.9*	-81.3*	-83.2*
Neutropenia 1	0	0	0	0	0	0.357	0.331	0.303	0.190	0.640	0.224	0.328
Neutropenia 2	70	0	0	20	0	-0.956	-1.26	-3.39	-0.764	-1.64	-1.56	-1.21
Thrombopenia 1	40	0	0	0	7.11*	-7.65*	-2.78	-2.39	-2.39	-2.81	-2.42	-2.62
Anemia 1	0	0	0	0	-3.09*	0.445	-0.0263	-0.0508	0.0188	0.0380	-0.0572	0.000632
Anemia 2	50	0	0	20	-38.3*	-5.99	-14.2*	-7.35	-1.93	-13.6*	-10.7*	-9.21*
<b>Number of selected variables</b>					6	3	3	1	2	3	3	3
<b>RMSE</b>					25.2	17.6	17.3	18.5	17.1	17.5	17.5	17.3
<b>-Log pointwise predictive density</b>					-	142	141	144	141	142	143	-

\* represents the variables selected by the model.

Abbreviations: BMA, Bayesian Model Averaging; SSVS, stochastic search variable selection; WBS, stochastic search variable selection; WHO, World Health Organization.

In our database, we found 185 patients who had data for the first  $K = 5$  cycles, among whom 70 patients had at least one cycle with complete toxicity data available. To test our method with a small sample, we randomly selected  $n = 10$  patients having  $N = 33$  complete forms. Among the 10 included patients, only 3 toxicities of grade 3 (asthenia) were observed (see Table 10 for description of covariates). For the case study, RMSEs were computed over these patients.

Results are shown in Table 11. The best performance was obtained by WBS3, with the smallest RMSE and two variables selected: age > 80 years and asthenia of grade 3 respectively linked to dose reductions of  $-13.4 \text{ mg/m}^2$  and  $-85.3 \text{ mg/m}^2$ . Asthenia 3 is selected by all methods, except the Lasso, and age is selected by these same methods, except WBS2. The clinical relevance weight linked to age is 100 for WBS1, WBS3, and WBS4 but is only 60 for WBS2. Only WBS2 yields worse performance than the classical SSVS, with only one variable selected. The Lasso selects six variables. WBS1, WBS4, and WBS with mixture and WBS with BMA choose the same variables. WBS3 and WBS4 select fewer variables than WBS1 and WBS4, possibly because of smaller sums of weights (cf Table 1). Other selected variables are anemia grade 2, selected by three methods (coefficient <  $-10$ ); asthenia grade 2 and thrombopenia grade 1, only selected by the Lasso and SSVS;

and WHO score and anemia grade 1, only selected by the Lasso. The WBS methods do not select thrombopenia grade 1, likely because this toxicity has clinical relevance weights of only 40 and 0.

In general, the WBS method produces coherent results, with WBS likely to correctly select ordinal variables having the highest clinical value, while the Lasso appears to select variables almost completely randomly, for instance, selecting asthenia grade 2 while not selecting asthenia grade 3.

## 6 | DISCUSSION

In this paper, we have proposed the WBS (weight-based SSVS) method derived from SSVS (stochastic search variable selection). The WBS method uses elicited clinical relevance weights to construct prior distributions for covariate coefficient inclusion probabilities in a regression model for longitudinal clinical practice data. An extensive simulation study showed that, as long as clinicians do not provide weights that make no sense, compared with the classical SSVS method, the WBS method exhibited better performance for all criteria considered (RMSE, log pointwise predictive density) and produced lower rates of both false positives and false negatives. As expected, performance improved with increasing sample size. WBS performance depended not only on the covariates' importance weights but also on the weights' sum, which must be calibrated carefully to obtain good performance. In our simulation study, we chose a large number of covariates implicated in dose reduction, and therefore, models with lower prior weights showed poorer performance because of low variable selection rates. The Lasso showed poor performance compared with WBS and SSVS, confirming that Bayesian methods outperform frequentist methods when working with small samples.

To our knowledge, variable selection methods incorporating informative prior distributions have been considered mainly in genomic settings. In this context, priors are established by exploiting an expansive literature on genetic variant severity, both from experimental and bioinformatic points of view. Therefore, in genetic settings, variability of priors is considered to be low, and simulation studies of performance variability due to prior specification have received little interest.<sup>19</sup> This is not the case for covariates retrieved from EHR, for which the literature is extremely limited regarding their respective effects on disease severity and thus their importance in medical decision making.

In our setting, only binary variables were used. We dichotomized continuous variables because it is more straightforward for clinicians to provide elicited weights for binary variables. The proposed method allows for the use of continuous variables, however. Another limitation of our formulation is that no hierarchical constraints were imposed in variable selection; that is, if one variable was included, nothing forced another to be included. For instance, if vomiting grade 3 is included in the model, vomiting of grade 4 should also be included. This limitation could be overcome by using the Farcomeni approach developed for SSVS models.<sup>14</sup> Additionally, in the Irinotecan data and similar settings, one may force all estimated regression coefficients to be negative or null because associated variables should imply either a dose reduction or no dose adjustment. Furthermore, this work began the analysis only from the second cycle, as we focused on toxicities. However, the first cycle can be included by considering all toxicities in this cycle to be null.

Many additional elaborations are possible. In this study, although the visit times for different patients may differ substantially, depending on the side effects generated in the previous cycle, we did not incorporate time or covariates of previous cycles. In our model, toxicities are not dependent variables but rather are covariates, and if there is a trend with respect to time, it should have already be summarized by the toxicities. In clinical practice, oncologists consider all of the patient's treatment history in choosing the dose and the next visit time. Furthermore, toxicities depend directly on previous doses and times. One possible extension might be to adapt this method to bivariate models in a dynamic treatment regime framework. Moreover, our model does not take into account the probability of changing the dose but only "how to reduce the dose when the clinician should to do it." Another extension of our model could be a conditional model on "when to change."

Concerning elicitation of clinical relevance weights, our case study results suggest that clinicians may not have an accurate perception of their actual decision process for dose adjustment. Indeed, most individuals are influenced by cognitive, psychological, and emotional factors that often prevent decision makers from choosing the best rational option available.<sup>26</sup> For example, in clinical practice, if a physician observes major side effects in a patient one day, his/her emotions and actual experience can lead him/her to reduce the dose of the next patient, even if common sense says to maintain the current dose.

The elicited scores differed between clinicians, as may be expected. When analyzing real data, one solution could be to combine the physicians' experiences to build more informative priors either by applying a mathematical aggregation rule, such as using a mixture prior with the physicians weighted equally, or by allowing the experts to interact with each other

to obtain a consensus prior.<sup>27,28</sup> This interaction may be face-to-face or may involve exchanges of information without direct contact. The prevailing mathematical approaches are averaging and pooling.<sup>29,30</sup> Other such approaches deviate from these traditional approaches by treating the elicited information as data. In this paper, we have proposed using BMA in order to combine the results obtained by the WBS method for each clinician's set of weights, giving higher importance to the best performing model.

Finally, modeling medical decision making may be regarded as a first step to modeling the complex relationships between covariates, dose reduction decisions, and survival. Our methodology may be extended to focus on dose combinations in frail patients, with the goal to optimize survival. Finally, our method may also be applied in other settings in which we wish to account for experts' opinion when selecting specific variables in small samples. Indeed, after choosing what variables to include in the model and categorizing them, clinical relevance weights may be elicited from many clinicians for each variable. Then, to combine the different sets of clinical weights, a mixture prior can be used after fitting the models for each clinician's weights or directly BMA as suggested in this paper. With this approach, the methodology is generally applicable to other settings having this data structure.

## ACKNOWLEDGEMENTS

The authors thank Dr Aziz Zaanan, Dr Céline Lepère, Dr Simon Pernot, and Dr Anne-Laure Pointet from Georges Pompidou European Hospital for their help with elicitation. The authors thank Angelika Geroldinger as well for her useful suggestions. The simulations were performed at the HPCaVe at UPMC-Sorbonne Université. The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: A part of this project was supported by French National Cancer Institut (INCa) under grant INCA\_10801. Moreno Ursino was funded by INCa grant INCA\_9539.

## CONFLICT OF INTEREST

The authors declare no potential conflict of interests.

## FINANCIAL DISCLOSURE

None reported.

## AUTHOR CONTRIBUTIONS

Anne-Sophie Jannot and Sarah Zohar made equal contributions and are co-last authors.

## ORCID

Sandrine Boulet  <https://orcid.org/0000-0001-9863-5466>

Moreno Ursino  <https://orcid.org/0000-0002-5709-4322>

## REFERENCES

1. Frankovich J, Longhurst CA, Sutherland SM. Evidence-Based Medicine in the EMR Era. *N Engl J Med*. 2011;365(19):1758-1759.
2. Tenenbaum JD, Avillach P, Benham-Hutchins M, et al. An informatics research agenda to support precision medicine: seven key areas. *J Am Med Inform Assoc*. 2016;23(4):791-795.
3. Bailey P, Chang DK, Nones K, et al. Genomic analyses identify molecular subtypes of pancreatic cancer. *Nature*. 2016;531:47-52.
4. Tibshirani R. Regression shrinkage and selection via the lasso: a retrospective. *J Royal Stat Soc Ser B Stat Methodol*. 2011;73(3):273-282.
5. Fan J, Li R. Variable selection via nonconcave penalized likelihood and its oracle properties. *J Am Stat Assoc*. 2001;96(456):1348-1360.
6. Efron B, Hastie T, Johnstone I, Tibshirani R. Least angle regression. *Ann Stat*. 2004;32(2):407-451.
7. Zou H, Hastie T. Regularization and variable selection via the elastic net. *J Royal Stat Soc Ser B Stat Methodol*. 2005;67(2):301-320.
8. Bondell HD, Reich BJ. Simultaneous regression shrinkage, variable selection and clustering of predictors with OSCAR. *Biometrics*. 2008;64(1):115-123.
9. O'Hara RB, Sillanpää MJ. A review of Bayesian variable selection methods: what, how and which. *Bayesian Analysis*. 2009;4(1):85-117.
10. Mitchell TJ, Beauchamp JJ. Bayesian variable selection in linear regression. *J Am Stat Assoc*. 1988;83(404):1023-1032.
11. George EI, McCulloch RE. Variable selection via Gibbs sampling. *J Am Stat Assoc*. 1993;88(423):881-889.

12. Kuo L, Mallick B. Variable selection for regression models. *Sankhyā Indian J Stat Ser B (1960-2002)*. 1998;60(1):65-81.
13. Ishwaran H, Rao JS. Spike and slab variable selection: frequentist and Bayesian strategies. *Ann Stat*. 2005;33(2):730-773.
14. Farcomeni A. Bayesian constrained variable selection. *Stat Sin*. 2010;20:1043-1062.
15. Zhang L, Baladandayuthapani V, Mallick BK, et al. Bayesian hierarchical structured variable selection methods with application to molecular inversion probe studies in breast cancer. *J Royal Stat Soc Ser C Appl Stat*. 2014;63(4):595-620.
16. Hobert JP, Casella G. The effect of improper priors on Gibbs sampling in hierarchical linear mixed models. *J Am Stat Assoc*. 1996;91(436):1461-1473.
17. Park T, Casella G. The Bayesian lasso. *J Am Stat Assoc*. 2008;103(482):681-686.
18. Chipman H, George EI, McCulloch RE. The practical implementation of Bayesian model selection. *Inst Math Stat Lect Notes Monogr Ser*. 2001;38:65-116.
19. Kitchen CMR, Weiss RE, Liu G, Wrin T. HIV-1 viral fitness estimation using exchangeable on subsets priors and prior model selection. *Statist Med*. 2007;26(5):975-990.
20. Thall PF, Ursino M, Baudouin V, Alberti C, Zohar S. Bayesian treatment comparison using parametric mixture priors computed from elicited histograms. *Stat Methods Med Res*. 2017;962280217726803.
21. Manceau G, Imbeaud S, Thiebaut R, et al. Hsa-miR-31-3p expression is linked to progression-free survival in patients with KRAS wild-type metastatic colorectal cancer treated with anti-EGFR Therapy. *Clin Cancer Res*. 2014;20(12):3338-3347.
22. Jansman FGA, Sleijfer DT, Coenen JLLM, De Graaf JC, Brouwers JRBJ. Risk factors determining chemotherapeutic toxicity in patients with advanced colorectal cancer. *Drug Safety*. 2000;23(4):255-278.
23. Bekele BN, Thall PF. Dose-finding based on multiple toxicities in a soft tissue sarcoma trial. *J Am Stat Assoc*. 2004;99(465):26-35.
24. Raftery AE, Madigan D, Hoeting JA. Bayesian model averaging for linear regression models. *J Am Stat Assoc*. 1997;92(437):179-191.
25. Gelman A, Hwang J, Vehtari A. Understanding predictive information criteria for Bayesian models. *Stat Comput*. 2014;24(6):997-1016.
26. Gorini A, Pravettoni G. An overview on cognitive aspects implicated in medical decisions. *Eur J Intern Med*. 2011;22(6):547-553.
27. Winkler RL. The consensus of subjective probability distributions. *Management Science*. 1968;15(2):B61.
28. Clemen RT, Winkler RL. Combining probability distributions from experts in risk analysis. *Risk Analysis*. 1999;19(2):187-203.
29. Genest C, McConway KJ. Allocating the weights in the linear opinion pool. *J Forecast*. 1990;9(1):53-73.
30. Genest C, Zidek JV. Combining probability distributions: a critique and an annotated bibliography. *Statistical Science*. 1986;1(1):114-135.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**How to cite this article:** Boulet S, Ursino M, Thall P, Jannot A-S, Zohar S. Bayesian variable selection based on clinical relevance weights in small sample studies—Application to colon cancer. *Statistics in Medicine*. 2019;1–20. <https://doi.org/10.1002/sim.8107>

## APPENDIX A

### CONVERGENCE OF RMSE

Take the simple case of the linear regression model

$$d_i = \theta_0 + \boldsymbol{\theta}^T \mathbf{z}_i + \epsilon_i, \quad \epsilon_i \text{ iid } \sim \mathcal{N}(0, \sigma^2), \quad i \in \{1, \dots, n\}.$$

Once coefficients  $(\theta_0, \boldsymbol{\theta}^T)$  have been estimated, we predict the dose  $d_i$  by

$$\hat{d}_i = \hat{\theta}_0 + \hat{\boldsymbol{\theta}}^T \mathbf{z}_i.$$

Because  $\text{MSE}(\hat{\mathbf{d}}) = \text{RMSE}(\hat{\mathbf{d}})^2 = \mathbb{E}((\mathbf{d} - \hat{\mathbf{d}})^2)$ , the MSE can be estimated by

$$\begin{aligned} \text{MSE} &= \frac{1}{n} \sum_{i=1}^n (d_i - \hat{d}_i)^2 \\ &= \frac{1}{n} \sum_{i=1}^n \left( (\theta_0 - \hat{\theta}_0) + (\boldsymbol{\theta}^T - \hat{\boldsymbol{\theta}}^T) \mathbf{z}_i + \epsilon_i \right)^2. \end{aligned}$$

In the ideal case in which  $(\hat{\theta}_0, \hat{\theta}^T) = (\theta_0, \theta^T)$ , we obtain

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n \epsilon_i^2 = \frac{\sigma^2}{n} \sum_{i=1}^n \left(\frac{\epsilon_i}{\sigma}\right)^2.$$

Now,  $\left(\frac{\epsilon_i}{\sigma}\right)_{1 \leq i \leq n}$  are independent, standard normal random variables; thus, the sum of their squares is distributed according to a chi-squared distribution with  $n$  degrees of freedom

$$\sum_{i=1}^n \left(\frac{\epsilon_i}{\sigma}\right)^2 \sim \chi_n^2.$$

We deduce the expectation and the variance

$$\mathbb{E}(\text{MSE}) = \sigma^2, \quad \mathbb{V}(\text{MSE}) = 2 \times \frac{\sigma^4}{n}.$$

## APPENDIX B

### PRACTICAL COMPUTATION OF LOG POINTWISE PREDICTIVE DENSITY

To compute the LPD in practice, it is possible to evaluate the expectation using draws from  $p_{\text{post}}(\theta, \mathbf{I}, \boldsymbol{\gamma}, \sigma_\gamma^2, \sigma_\epsilon^2)$ , the usual posterior simulations, which are labeled  $\theta^s, \mathbf{I}^s, \boldsymbol{\gamma}^s, c^{2s}, \sigma_\epsilon^{2s}, s = 1, \dots, S$ ,

$$\widehat{\text{LPD}} = \sum_{i=1}^n \sum_{k=1}^K \log \left( \frac{1}{S} \sum_{s=1}^S p(d_{i,k} | \theta^s, \mathbf{I}^s, \boldsymbol{\gamma}^s, c^{2s}, \sigma_\epsilon^{2s}) \right).$$

The log pointwise predictive density  $\widehat{\text{LPD}}$  is the sum over patients and cycles of the log of the mean over MCMC iterations of the probability that a new dose  $\tilde{d}$  would be obtained by the estimated model.

The following steps are suggested to compute the LPD in our case:

1. For  $s = 1, \dots, S$ , sample  $(\theta^s, c^{2s}, \sigma_\epsilon^{2s})$  from  $p_{\text{post}}(\theta, \boldsymbol{\gamma}, \sigma_\gamma^2, \sigma_\epsilon^2)$ :  
 $(\theta^s, c^{2s})$  are the values obtained for the  $s$ th iteration of the MCMC;
2. For  $s = 1, \dots, S$ , for  $i = 1, \dots, n$ , draw  $\gamma_i^s$  from  $\mathcal{N}(0, c^{2s})$ ;
3. Compute the probability that  $p(\tilde{\mathbf{d}} | \theta^s) = (\sigma_\epsilon^{2s})^{-\frac{nk}{2}} \exp \left( -\frac{(\tilde{\mathbf{d}} - \theta_0^s - \mathbf{z}\theta^s - \boldsymbol{\gamma}^s)^T (\tilde{\mathbf{d}} - \theta_0^s - \mathbf{z}\theta^s - \boldsymbol{\gamma}^s)}{2\sigma_\epsilon^{2s}} \right)$ .



# Intégration de l'élicitation d'experts dans une méthode de sélection de variables en Bayésien par la méthode du *power prior* - Application au cancer du colon

**Contexte :** Construire des outils d'aide à la décision en médecine nécessite d'identifier les variables pertinentes pour modéliser la décision médicale. À cette fin, les données expertes et les données patients constituent toutes deux des sources importantes d'information. Actuellement, des techniques d'apprentissage automatique permettent de sélectionner les variables pertinentes utilisées pour la prise de décision à partir des données de soin de patients. Mais aucune méthode ne permet d'utiliser de manière conjointe données de soin et données expertes. Nous proposons une méthode qui introduit de l'information experte dans un modèle de sélection de variables. Nous nous plaçons dans le contexte des décisions médicales relatives à l'adaptation des doses d'irinotécan pour le traitement du cancer colorectal métastatique.

**Méthodes :** Nous avons extrait les données de soin des patients ayant eu un protocole incluant de l'irinotécan entre 2012 et 2017 à partir des dossiers patients informatisés de l'Hôpital Européen Georges Pompidou. Le service d'oncologie de cet hôpital a mis en place depuis 2012 un questionnaire standardisé collectant l'ensemble des informations nécessaires pour les adaptations de dose de chimiothérapie. Les données expertes proviennent des poids estimés par les experts associés aux variables impliquées dans les décisions de réduction de dose : nous avons demandé aux cliniciens experts d'estimer sur une échelle allant de 0 à 100 quels poids ils accordaient à chaque variable utilisée pour les adaptations de dose d'irinotécan. Un échantillon de doses simulées a ensuite été généré à partir de ces poids élicités puis combiné aux données de réductions de doses observées via la méthode du *power prior* [50]. Nous avons modélisé le lien entre réductions de dose répétées, caractéristiques des patients et toxicités par le biais d'un modèle logistique à effets mixtes. La sélection de variables a ensuite été réalisée via le modèle *Stochastic Search Variable Selection* (SSVS) [38]. Les performances de cette méthode combinant information observée et information experte ont été comparées à celles de la méthode utilisant uniquement l'information observée pour notre cas d'usage, et une analyse de sensibilité basée sur le paramètre de *power prior* a été réalisée.

**Résultats :** Les poids élicités par les experts sont très variables. Les performances prédictives des différents modèles sont modérées avec un AUC proche de 0,7. Notre méthode permet de sélectionner des variables rares qui pourraient être omises sur la base des données observées seules et d'éliminer des



variables qui semblent pertinentes à partir des données, mais avec des poids élicités faibles.

**Conclusion :** Nous présentons une méthode Bayésienne de sélection de variables incorporant l'information experte élicitée et les données réelles. La méthode sélectionne un ensemble de variables pertinentes pour modéliser le processus de décision médicale. Les poids élicités sont très variables selon les cliniciens ce qui souligne la variabilité des pratiques professionnelles.

# Integration of elicited expert information via a power prior in Bayesian variable selection: Application to colon cancer data

Statistical Methods in Medical Research  
0(0) 1–27

© The Author(s) 2019

Article reuse guidelines:

[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)

DOI: 10.1177/0962280219841082

[journals.sagepub.com/home/smm](https://journals.sagepub.com/home/smm)

Sandrine Boulet,<sup>1</sup> Moreno Ursino,<sup>1</sup> Peter Thall,<sup>2</sup> Bruno Landi,<sup>3</sup>  
Céline Lepère,<sup>3</sup> Simon Pernot,<sup>3</sup> Anita Burgun,<sup>1,4</sup> Julien Taieb,<sup>3</sup> Aziz Zaanan,<sup>3</sup>  
Sarah Zohar<sup>1</sup> and Anne-Sophie Jannot<sup>1,4</sup>

## Abstract

**Background:** Building tools to support personalized medicine needs to model medical decision-making. For this purpose, both expert and real world data provide a rich source of information. Currently, machine learning techniques are developing to select relevant variables for decision-making. Rather than using data-driven analysis alone, eliciting prior information from physicians related to their medical decision-making processes can be useful in variable selection. Our framework is electronic health records data on repeated dose adjustment of Irinotecan for the treatment of metastatic colorectal cancer. We propose a method that incorporates elicited expert weights associated with variables involved in dose reduction decisions into the Stochastic Search Variable Selection (SSVS), a Bayesian variable selection method, by using a power prior.

**Methods:** Clinician experts were first asked to provide numerical clinical relevance weights to express their beliefs about the importance of each variable in their medical decision making. Then, we modeled the link between repeated dose reduction, patient characteristics, and toxicities by assuming a logistic mixed-effects model. Simulated data were generated based on the elicited weights and combined with the observed dose reduction data via a power prior. We compared the Bayesian power prior-based SSVS performance to the usual SSVS in our case study, including a sensitivity analysis using the power prior parameter.

**Results:** The selected variables differ when using only expert knowledge, only the usual SSVS, or combining both. Our method enables one to select rare variables that may be missed using only the observed data and to discard variables that appear to be relevant based on the data but not relevant from the expert perspective.

**Conclusion:** We introduce an innovative Bayesian variable selection method that adaptively combines elicited expert information and real world data. The method selects a set of variables relevant to model medical decision process.

## Keywords

Bayesian variable selection, clinical relevance weights elicitation, power prior method, repeated measures, electronic health record

## 1 Introduction

Electronic health records (EHR) are used increasingly in hospitals.<sup>1,2</sup> With the development of precision/personalized medicine, medical decision-making has become an increasingly complex process and the use of

<sup>1</sup>INSERM U1138, University Paris Descartes, Sorbonne University, Paris, France

<sup>2</sup>Department of Biostatistics, M.D. Anderson Cancer Center, Houston, Texas, USA

<sup>3</sup>Department of Digestive Oncology, Hôpital européen Georges-Pompidou, AP-HP, Paris, France

<sup>4</sup>Department of Statistics, Medical Informatic and Public Health, Hôpital européen Georges-Pompidou, AP-HP, Paris, France

### Corresponding author:

Sandrine Boulet, INSERM U1138, University Paris Descartes, Sorbonne University, Paris, France.

Email: [sandrine.boulet@crc.jussieu.fr](mailto:sandrine.boulet@crc.jussieu.fr)

existing routine data has become essential to design Clinical Decision Support Systems (CDSS). Until now, most CDSS have been knowledge-based systems (also called expert systems since they emulate the decision-making ability of a medical expert). These systems provide explicit algorithms that incorporate reasoning capabilities based on inference engines and knowledge bases, including validated rules, shared guidelines, and ontologies. However, in routine care, medical decision is a quite complex process and all decisions cannot be modelled using knowledge-based models. According to Smellie et al., available guidelines are mostly focused on a particular disease and provide (sometimes very) limited advice for specific patient-centric interpretation.<sup>3</sup> Knowledge-based models might not cover all types of situations. In real world, physicians may optimize their medical decision-making process based on multiple patients characteristics and their experience, as they can rely only very partially on guidelines. Therefore, some variables might be taken into account in routine care, while not appearing in knowledge-based models. Consequently, algorithms based on high dimensional routine data analyses are developing in health care, thanks to the increasing performance of data mining and machine learning techniques. For instance, it was already proposed to combine data mining and case-based reasoning for intelligent decision support for pathology ordering by general practitioners.<sup>4</sup> To apply such methods on routine care data, we have to hypothesize that the variables that influence clinical decisions can be retrieved from the EHR. In selecting relevant variables for use in CDSS, analysis of EHR data may be highly informative. Nevertheless, such routine data analyses also may reflect cognitive bias, such as discrimination.<sup>5</sup> Moreover, machine learning algorithm-based on EHR data alone may miss rare but important variables, due to a lack of power.

Rather than using either data-driven analysis alone or only expert knowledge to build an algorithm for supporting decision making, a promising alternative is to combine these two types of information. Expert opinion can be incorporated into routine practice data by eliciting from physicians a numerical importance weight of each variable used in their decision making. Thus, if physicians assign a very low weight to some variable, it should not be selected. Concomitantly, if the number of observations does not allow us to detect the effect of a rare variable using only the observed data, then this elicitation could make it possible to reinforce this signal and thus allow the variable to be selected. In this paper, we propose to combine information from EHR data and elicited expert variable importance weights to select a minimal set of relevant variables for use in dose and treatment decisions in the context of multistage chemotherapy.

Building personalized decision support tools to adapt doses in anti-cancer drug regimen is of particular interest since chemotherapy may involve many decisions regarding the choice of drugs and doses. Numerous variables (pharmacogenetic markers, tumor mutations, clinical characteristics) may be involved in the choice of an anti-cancer treatment protocol. Usually, each patient receives his/her first protocol combination dose according to his/her weight, height, age, and biomarkers. During subsequent treatment cycles, variables such as weight loss, ECOG (Eastern Cooperative Oncology Group) performance status, treatment line, response, pharmacogenetic polymorphisms, and toxicities can influence the physicians' medical decision to either reduce the dose or choose a different treatment line (protocol). Post-hoc analysis of randomized clinical trials and real world data have already highlighted a high frequency of dose adjustment according to patients' advanced age, and comorbidities such as diabetes, obesity.<sup>6-8</sup> Moreover, some variables may have a known link to certain toxicities.<sup>9</sup> Therefore, physicians must adapt doses and schedules over consecutive treatment cycles by considering all characteristics of the patient and past treatment lines, doses, and outcomes. Thus, decisions are personalized, and usually only a few observations for a given set of characteristics are available in routine data. Consequently, using data-driven strategies, rare variables that are very important for medical decisions might not be selected.

Because of the large number of variables in an EHR dataset, to select variables relevant for medical decision making, one can use penalization criteria such as bridge, Lasso, SCAD, LARS, elastic net, or OSCAR regressions.<sup>10-14</sup> In the Bayesian setting, Mitchell and Beauchamp have suggested a method based on "spike and slab" priors,<sup>15</sup> with its extension, the Stochastic Search Variable Selection (SSVS) method, that use a mixture of two normal distributions centered on zero, but with different variances as proposed by George and McCulloch.<sup>16</sup> The Bayesian approach can also enable us to incorporate prior expert information.

To combine information from EHR data and expert opinion, we propose using a power prior. Traditionally, power prior methods are used to combine historical data with current data via an informative prior, called the power prior, using Bayesian regression.<sup>17</sup> The power prior parameter allows one to tune how much historical information should be used in posterior estimation of the model parameters. If the power prior parameter is equal to zero, then no historical information is used, and the prior distribution is taken as non informative. If the power prior parameter is equal to one, then all the informative prior distribution is used to compute the posteriors.

Our motivating application is based on observed data derived from an EHR of colorectal cancer patients treated with Irinotecan. Our expert data were obtained from six digestive oncologists by quantifying their subjective importance of variables used in their own decision making. In our proposed method, historical data are replaced by simulated data. The general steps of how we applied the methodology to our motivating application are as follows:

- (1) First, we independently asked six clinicians who are experts in treating colorectal cancer to specify covariates and numerical clinical relevance weights to reflect their opinions about the importance of each covariate in their decision making. When a covariate takes on a particular value, a clinician will decrease the dose with a probability that relies on the expert's clinical relevance weights given to this variable. Each clinician gave their own clinical relevance weights linked to each grade of each toxicity type and each level of each covariate.
- (2) Next, we formulated a Bayesian model for the observed data  $\mathbf{D}$ , including a spike-and-slab prior to facilitate variable selection using SSVS.
- (3) We generated a simulated data set  $\mathbf{D}_0$  using the clinical relevance weights, and we formulated a synthetic likelihood  $L(\theta|\mathbf{D}_0)$ .
- (4) We used a power prior to combine the usual noninformative prior  $\pi_0(\theta)$  with the synthetic likelihood  $L(\theta|\mathbf{D}_0)$  to obtain the prior we actually used. The proposed method combines the physician weights in the posterior as a function of estimated values from the observed data and the elicited weights.

The remainder of the article is organized as follows. In section 2, we describe the case study of patients treated with Irinotecan for colorectal cancer. The proposed methodology is presented formally in section 3. We present the results of our application of the method to the clinical practice data in section 4, and we discuss them in section 5. We close with conclusions in section 6.

## 2 Case study

To ensure complete reporting of our routinely collected health data, we followed the RECORD statement.<sup>18</sup>

### 2.1 Patient selection

The chemotherapy data analyzed herein were extracted from the patient EHR and stored in an i2b2 clinical data warehouse together with all other hospital health records.<sup>19,20</sup> These data warehouses have been established to facilitate health care data reuse for research. We selected data from the Georges Pompidou University Hospital (HEGP) i2b2 data warehouse, on patients treated for metastatic colorectal cancer who received at least once a combination of drugs, including Irinotecan at a theoretical dose of 180 mg/m<sup>2</sup> and a theoretical chemotherapy cycle of 14 days, between the 1 July 2003 and 10 November 2017. Selected patients were then split into two samples, one (hereafter called sample 1) with patients having at least one cycle with complete information, and the other (hereafter called sample 2) with patients not fulfilling this criterion.

### 2.2 Cycle selection

For sample 1, if the patient experienced a dose reduction (dose less than 150 mg/m<sup>2</sup>) on a cycle with complete observation, we extracted the maximal number of consecutive cycles that included this cycle. Otherwise, we extracted the maximal number of consecutive cycles with complete observation. This procedure offers the opportunity to enrich our sample of patients with dose reductions.

For sample 2, we kept the maximal number of consecutive cycles with less than  $u$  missing data by cycle for  $u \in \{1, \dots, 10\}$ .

### 2.3 Variable extraction

Chemotherapy protocols and prescriptions are defined in the Chimio<sup>®</sup> (Computer Engineering) software that is used by clinicians for prescription. This software manages the dates of administration for each cycle, the doses of the different drugs, and dose reductions, among others. Some clinical data are also stored by clinicians in Chimio<sup>®</sup>, such as body surface area or creatinine clearance.

**Table 1.** Definitions of neutropenia, thrombopenia, and anemia.

Toxicities		Grades			
		1	2	3	4
Neutropenia	Neutrophils (/mm <sup>3</sup> )	<2000	<1500	<1000	<500
Thrombopenia	Blood platelets (/mm <sup>3</sup> )	<150,000	<75,000	<50,000	<25,000
Anemia	Hemoglobin (g/dL)	<12	<10	<8	<6.5

The screens of the Chimio<sup>®</sup> application were reviewed to determine the relevant items to upload into the i2b2 data warehouse.

A total of 1.4 million observations were included from Chimio<sup>®</sup> in i2b2, involving 14,000 patients. 1.300 i2b2 concepts were created to store these observations.

To obtain clinical data associated with each chemotherapy cycle, a questionnaire was designed in 2012 at the HEGP to collect the clinical information needed for patient follow-up and dose adjustments. In particular, it contains the toxicities of chemotherapies at each cycle, by organ, graded for severity according to the Common Terminology Criteria for Adverse Events (CTCAE) v4.03 nomenclature. We extracted these toxicities along with other relevant variables (age, ECOG performance status, bilirubin, weight, treatment line) and included laboratory values (neutrophils, platelets, hemoglobin) available in our clinical data warehouse (cf. Table 1).

If several clinical reports were available for one cycle, we used the maximal grade for each toxicity variable.

We derived the following relevant variables from the clinical report:

- Age < or  $\geq 80$  years;
- Weight loss < or  $\geq 10\%$ , since the start of treatment;
- Bilirubin categorized as < 35  $\mu\text{mol/L}$ , between 35 and 50  $\mu\text{mol/L}$ , or  $\geq 50 \mu\text{mol/L}$ ;
- Treatment line < 3, = 3 or > 3;
- ECOG performance status and toxicities graduated integer values from 0 (no side effect) to either 3 or 4 (most severe).

## 2.4 Weight elicitation

Let  $n$  be the number of patients and  $K_i$  be the number of cycles for patient  $i \in \{1, \dots, n\}$ . For each  $i$  and  $k \in \{1, \dots, K_i\}$ , let  $\mathbf{x}_{i,k}$  be the  $J$ -vector of variables at the beginning of cycle  $k$ . Let  $x'_{j,c}$  be the value for the  $c$ th category of the  $j$ th variable,  $j \in \{1, \dots, J\}$ ,  $c \in \{0, \dots, C_j\}$  where  $C_j + 1$  is the number of categories of this variable. Let  $\mathbf{z}_{i,k}$  be the  $L$ -vector ( $L = \sum_{j=1}^J C_j$ ) of all individual variables in  $\mathbf{x}_{i,k}$ , where each  $x_{i,k,j}$  is split into  $C_j$  dummy variables, each taking a value of 1 if the corresponding category was in  $x_{i,k,j}$  and 0 otherwise.

For each variable  $j \in \{1, \dots, J\}$  and each of its possible values  $c \in \{1, \dots, C_j\}$ , each physician was asked to specify the clinical relevance weights  $\mathbf{w}_j = (w_{j,1}, \dots, w_{j,C_j})$ . These weights quantify the importance of variables in dose adaptation decision-making based on individual clinician experience. The minimal and maximal possible values for the weights were restricted to 0 (no severity) and  $W_{max} = 100$  (the highest possible severity), since this scale is easily understandable by oncologists, and ordered as follows:  $0 = w_{j,0} \leq w_{j,1} \leq \dots \leq w_{j,C_j} \leq W_{max} = 100$ , such that lower grades are less involved in the dose reduction decision than higher grades. For each variable  $j$ ,  $w_{j,0} = 0$  is the weight for the reference level  $x'_{j,0}$ . Similarly, for  $c \in \{1, \dots, C_j\}$ ,  $w_{j,c}$  is the weight used when the variable  $x_{i,k,j}$  takes the value  $x'_{j,c}$ . For instance, the vector  $\mathbf{w}_7 = (0, 20, 80)$  is the vector of weights for the 7th variable at its  $C_7 = 3$  levels (the reference level is not included). By analogy with  $\mathbf{z}$ ,  $\mathbf{w}^*$  denotes the  $L$ -vector of all clinical relevance weights. This notation follows Bekele and Thall,<sup>21</sup> in which elicited toxicity severity weights were used to build a score called the "total toxicity burden" for use in adaptive dose-finding.

Table 2 presents the clinical relevance weights elicited from each clinician for each covariate. For example,  $\{0, 20, 80\}$  in the first column (physician 1), in the row "nausea" refers to clinical relevance weights given by physician 1 for grades  $\{1, 2, 3\}$ . The table suggests that the clinician 2 assigned a weight of 20 to a weight loss of more than 10% in his/her decision making, while clinician 4 assigned a weight of 80.

**Table 2.** Clinical relevance weights for each covariate elicited from each physician.

Variables	Physicians											
	1				2				3			
Age $\geq 80$ years	100	–	–	–	60	–	–	–	100	–	–	–
Weight loss $> 10\%$	50	–	–	–	20	–	–	–	50	–	–	–
ECOG performance status (1,2,3,4)	0	20	20	20	0	0	40	100	0	0	80	100
Bilirubin $> 35, > 50 \mu\text{mol/L}$	100	100	–	–	40	80	–	–	100	100	–	–
Treatment line 3, $> 3$	30	50	–	–	0	0	–	–	0	0	–	–
Toxicity grades 1, 2, 3, 4												
Vomiting	0	20	80	90	0	30	70	100	0	10	10	10
Nausea	0	20	80	–	0	10	50	–	0	10	10	–
Diarrhea	0	40	80	100	0	20	50	100	0	50	80	100
Asthenia	10	50	100	–	10	10	40	–	0	0	70	–
Neutropenia	0	70	100	100	0	0	30	50	0	0	50	50
Thrombopenia	40	100	100	100	0	0	20	30	0	0	50	50
Anemia	0	50	80	100	0	0	20	30	0	0	0	0

Variables	Physicians											
	4				5				6			
Age $\geq 80$ years	80	–	–	–	100	–	–	–	100	–	–	–
Weight loss $> 10\%$	80	–	–	–	50	–	–	–	60	–	–	–
ECOG performance status (1,2,3,4)	0	20	80	100	0	30	100	100	0	70	100	100
Bilirubin $> 35, > 50 \mu\text{mol/L}$	20	80	–	–	100	100	–	–	100	100	–	–
Treatment line 3, $> 3$	0	0	–	–	0	0	–	–	0	50	–	–
Toxicity grades 1, 2, 3, 4												
Vomiting	10	20	80	100	0	30	70	100	0	0	70	100
Nausea	10	30	80	–	0	10	40	–	0	0	30	–
Diarrhea	0	20	70	90	0	20	50	100	0	50	100	100
Asthenia	10	50	70	–	0	20	50	–	0	40	80	–
Neutropenia	0	20	70	80	0	0	20	100	0	0	0	50
Thrombopenia	0	50	80	100	0	0	20	70	0	0	0	0
Anemia	0	20	50	70	0	0	30	80	0	0	0	0

### 3 Methods

In the following section, we introduce the Bayesian mixed model framework, explain how we integrate the elicited expert weights into the statistical model, and describe criteria to assess predictive performance and variable selection.

#### 3.1 Bayesian mixed model framework

We use a mixed-effects logistic regression model where the dose reduction for patient  $i$  in cycle  $k$  is denoted by the binary variable  $r_{i,k}$ , and the covariate vector  $\mathbf{z}_{i,k}$  includes patient characteristics and toxicities, updated from the previous cycle

$$r_{i,k} \sim \mathcal{B}(p_{i,k}^r)$$

$$\text{logit}(p_{i,k}^r) = \theta_0 + \boldsymbol{\theta}^T \mathbf{z}_{i,k} + \gamma_i$$

$$\gamma_i \stackrel{iid}{\sim} \mathcal{N}(0, \sigma_\gamma^2)$$

where  $(\theta_0, \boldsymbol{\theta})$  is the  $L$ -vector of fixed coefficients,  $\gamma_i$  is the random patient effect, and  $\mathcal{B}(p^r)$  represents the Bernoulli distribution with a probability of success equal to  $p^r$ .

For  $l \in \{1, \dots, L\}$ , let  $I_l$  to be the selection indicator variable as follows

$$I_l = \begin{cases} 1 & \text{if } z_{\cdot,l} \text{ is in the model} \\ 0 & \text{otherwise} \end{cases}$$

where  $z_{\cdot,l}$  is the  $l$ th variable. We perform variable selection using the Stochastic Search Variable Selection (SSVS) method, which assumes that  $\theta_l|I_l$ ,  $l \in \{1, \dots, L\}$  follows a mixture of two normal distributions, both centered on 0 but with different variances: the first variance is chosen to be small so that unselected variables have estimated coefficients that are very likely to be close to 0, and the second is chosen to be large so that estimated coefficients linked to selected variables can assume a wide range of real values with nontrivial probability.<sup>16</sup> Formally, the distribution of  $\theta_l|I_l$ ,  $l \in \{1, \dots, L\}$  is assumed to be

$$\theta_l|I_l \sim (1 - I_l)TN(0, \tau^2) + I_lTN(0, g\tau^2) \quad (1)$$

where  $(\tau, g)$  are variance parameters, the selection variable indicator  $I_l$  follows a Bernoulli distribution with parameter  $p_l$ , and  $TN$  is the truncated normal distribution on  $(-\infty, 0]$ . The truncated normal distribution restricts parameters to negative values, such that the model only accounts for dose reductions. In practice, doses specified per protocol are possibly decreased but not increased. Therefore, there were no dose increases in our data.

### 3.2 Integration of expert elicitation weights into the power prior method

In this section, we describe how we combined the elicited expert weights with the observed data when modeling dose reductions.

The power prior (PP) method provides a basis for integrating information contained in historical data into a Bayesian analysis,<sup>17</sup> while calibrating the weight of the historical data relative to the current data. In our proposed method, historical data are replaced by simulated data generated using patients with incomplete data (sample 2). For this dataset, we first (1) imputed missing values at random according to the variable distribution in the initial data set and (2) generated dose reductions  $r_0$  according to the elicited weights.

Two alternative approaches were used to generate simulated dose reductions  $r_0$  according to the experts' elicited weights:

- Approach 1:

$$r_{0,i',k'}^{(1)} = \begin{cases} 1 & \text{if } \mathbf{w}^{*T} \mathbf{z}_{0,i',k'} > Q_0 \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

- Approach 2:

$$\begin{aligned} r_{0,i',k'}^{(2)} &\sim \mathcal{B}(p_{0,i',k'}^r) \\ \text{logit}(p_{0,i',k'}^r) &= -Q_0 + \mathbf{w}^{*T} \mathbf{z}_{0,i',k'} \end{aligned} \quad (3)$$

In both approaches,  $Q_0$  is a threshold defined to obtain a proportion  $P_0$  of dose reductions, in which  $Q_0$  is the quantile corresponding to the probability  $1 - P_0$  of the sample made up of  $(\mathbf{w}^{*T} \mathbf{z}_{0,i',k'})_{i',k'}$ . Equations (2) and (3) show how the elicited weights were used as covariate parameters in the synthetic data simulation model. Furthermore, the simulated data set depends on the  $P_0$  value that is chosen based on a sensitivity analysis, described in section 3.4. The second approach introduces uncertainty into the dose reduction and makes the process of decision making more similar to that of clinicians in clinical practice by sometimes discarding important variables.

The simulated data are combined with observed data using the PP. This prior is defined as the product of the weighted likelihood of  $\theta$  conditional on the simulated data and a noninformative prior distribution

$$\pi(\theta|\mathbf{D}_0, a_0) \propto L(\theta|\mathbf{D}_0)^{a_0} \pi_0(\theta)$$

where  $\mathbf{D}_0 = (n_0, \mathbf{r}_0, \mathbf{z}_0)$  are the simulated data,  $L(\theta|\mathbf{D}_0)$  is the likelihood of  $\theta$  conditional on the simulated data,  $0 \leq a_0 \leq 1$  is the scale parameter that controls the amount of weight assigned to the simulated data, and  $\pi_0(\theta)$  is a noninformative prior for  $\theta$ .



The posterior distribution of  $\theta$  is

$$\begin{aligned}\pi(\theta|\mathbf{D}, \mathbf{D}_0, a_0) &\propto L(\theta|\mathbf{D})\pi(\theta|\mathbf{D}_0, a_0) \\ &\propto L(\theta|\mathbf{D})L(\theta|\mathbf{D}_0)^{a_0}\pi_0(\theta)\end{aligned}$$

where  $\mathbf{D} = (n, r, z)$  are the observed data and  $L(\theta|\mathbf{D})$  is the likelihood of  $\theta$  conditional on the observed data.

### 3.3 Bayesian model averaging (BMA)

To combine the weights elicited from different physicians,  $c \in \{1, \dots, C\}$ , we use Bayesian Model Averaging (BMA) with the predictive distribution of a coefficient  $\theta_l$ ,  $l \in \{1, \dots, L\}$ , which is a weighted average of the posterior distributions of  $\theta_l$  using our model  $M_c$  for each clinician  $c$ <sup>22</sup>

$$\mathbb{P}(\theta_l|\mathbf{D}) = \sum_{c=1}^C \mathbb{P}(M_c|\mathbf{D})\mathbb{P}(\theta_l|M_c, \mathbf{D})$$

where the weights  $\mathbb{P}(M_c|\mathbf{D})$  are the posterior probabilities of each model

$$\mathbb{P}(M_c|\mathbf{D}) = \frac{\mathbb{P}(\mathbf{D}|M_c)\mathbb{P}(M_c)}{\sum_{c_1=1}^C \mathbb{P}(\mathbf{D}|M_{c_1})\mathbb{P}(M_{c_1})}$$

The marginal likelihood of the model  $M_c$  is

$$\mathbb{P}(\mathbf{D}|M_c) = \int \mathbb{P}(\mathbf{D}|\theta_l, M_c)\mathbb{P}(\theta_l|M_c)d\theta_l$$

where  $\mathbb{P}(\mathbf{D}|\theta_l, M_c)$  is the likelihood and  $P(\theta_l|M_c)$  is the prior distribution. The expected value and variance of  $\theta_l$  are given by

$$\begin{cases} \mathbb{E}(\theta_l|\mathbf{D}) = \sum_{c=1}^C \hat{\theta}_l^c \mathbb{P}(M_c|\mathbf{D}) \\ \mathbb{V}(\theta_l|\mathbf{D}) = \sum_{c=1}^C [\mathbb{V}(\theta_l|\mathbf{D}, M_c) + (\hat{\theta}_l^c)^2] \mathbb{P}(M_c|\mathbf{D}) - E(\theta_l|\mathbf{D})^2 \end{cases}$$

where  $\hat{\theta}_l^c = \mathbb{E}(\theta_l|\mathbf{D}, M_c)$ .

### 3.4 Prior parameter settings

Numerical values of prior hyperparameters  $g$  and  $\tau^2$  in equation (1) must be specified so that if  $\theta_l \sim \mathcal{N}(0, \tau^2)$ ,  $\theta_l$  then estimate  $\hat{\theta}_l$  will be close to 0, and if  $\theta_l \sim \mathcal{N}(0, g\tau^2)$ , then  $\hat{\theta}_l$  may take on nonzero real values. In the following, based on a preliminary sensitivity analysis, we chose  $\tau = 1$  and  $g = 100$  following the proposal in George and McCulloch.<sup>16</sup> We also took  $p_l = 0.5$ , so that all variables have the same probability of being selected by the algorithm. The intercept  $\theta_0$  and the covariance of the random effects  $\sigma_y^2$  were assumed to follow a normal distribution with a mean of 0 and variance of 100, and an inverse-gamma distribution with shape and scale parameters equal to 1 and 1, respectively.

For the PP, we performed a sensitivity analysis of prediction performance and variable selection. Two alternative approaches were used to generate the simulated dose reductions  $r_0$  according to the elicited expert weights. We considered the power prior parameter values ( $a_0 \in \{0.1, 0.5, 1\}$ ), simulated sample sizes ( $n_0 = n_1$ ,  $n_0 = 1.77 \times n_1$ ), and dose reduction frequencies ( $P_0 \in \{2P, P, P/2\}$ ).

### 3.5 Assessment of predictive performance and variable selection

To assess the performance and variable selection, we randomly split the sample with complete observations into two subsets, one for training (70%) and one for performance estimation (30%), and repeated this 10 times,



**Table 3.** Covariates for patients with complete observations over all cycles (sample 1).

Variables	0	1	2	3	4
Age $\geq 80$ years	604 (95)	32 (5)	–	–	–
Weight loss $> 10\%$	631 (99)	5 (1)	–	–	–
ECOG performance status	123 (19)	381 (60)	100 (16)	11 (2)	21 (3)
Bilirubin $> 35, > 50 \mu\text{mol/L}$	630 (99)	1 (0)	5 (1)	–	–
Treatment line 3, $> 3$	547 (86)	46 (7)	43 (7)	–	–
Toxicity grades					
Vomiting	567 (89)	54 (8)	13 (2)	2 (0)	–
Nausea	344 (54)	226 (36)	60 (9)	6 (1)	–
Diarrhea	363 (57)	218 (34)	49 (8)	6 (1)	–
Asthenia	97 (15)	368 (58)	150 (24)	21 (3)	–
Neutropenia	444 (70)	111 (17)	67 (11)	9 (1)	5 (1)
Thrombopenia	524 (82)	104 (16)	6 (1)	1 (0)	1 (0)
Anemia	343 (54)	257 (40)	36 (6)	–	–

Note: Percentages are shown in parentheses.

resulting in 10 use cases. Parameters were estimated for each training sample considered as observed data in the likelihood, with variable selection performed on the training sample. For each variable, we counted the proportion of times this variable was selected. Predictive performance then was estimated for each test sample. To assess the model predictive performance, we used the Receiver Operating Characteristic (ROC) curve and its corresponding Area Under the Curve (AUC).

We performed the analyses using *R* software 3.3.2 version and package *R2jags* (0.5-7 version). In *R2jags*, three chains and a burn-in of 1000 were used to implement *autojags*, combined to Gelman and Rubin's potential scale reduction factor  $Rhat = 1.1$  as a convergence criterion. The scripts are available from the corresponding author.

## 4 Results

Among the 678 patients in the sample,  $n = 303$  had complete observations over all cycles (sample 1; see Table 3, Figures 1 to 3 for descriptions). There were very few observations with grade 3 or 4 toxicities. Among patients with complete observations over all their cycles, 85 had dose reductions, corresponding to 13% of the observed dose reductions in sample 1.

This sample of 303 patients was used as observed data in the method. Thus, 10 training samples of  $n_1 = 212$  patients and 10 validation samples of 91 patients were built to assess the predictive performance of the proposed model. The 375 patients with incomplete data (sample 2) from the initial data set were used to build the simulated dataset by (1) imputing their missing values at random according to the variable distribution in the initial data set and (2) generating dose reductions according to the elicited weights (see Appendix 2 for description of this sample). A flowchart summarizing the use of the data is presented in Figure 4.

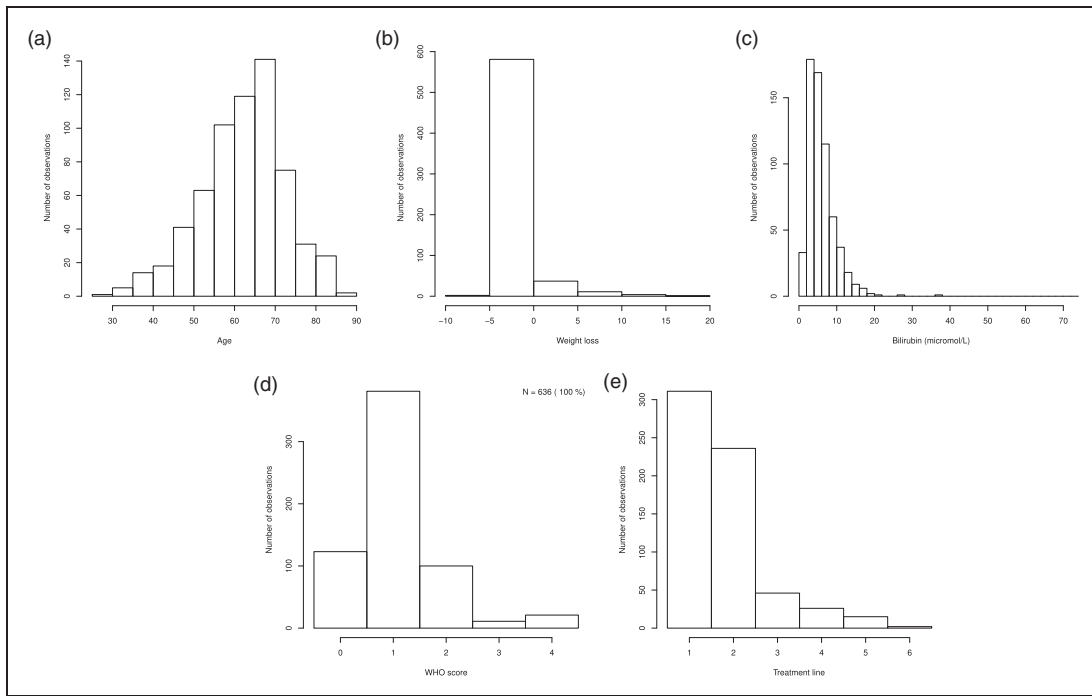
The elicitation weights showed that important variables for decision making were the ECOG performance status, except for physician 1, and grade 4 toxicities. Some variables were considered important less consistently, such as the treatment line or anemia. The elicited weights suggest that the six clinicians might act differently when dealing with the same clinical situation.

### 4.1 Prediction performance

Let  $P$  be the proportion of observed dose reductions in each of the 10 training datasets of sample 1. Table 4 summarizes the number of selected variables and AUCs for usual SSVS and SSVS using the PP model with  $a_0 \in \{0.1, 0.5, 1\}$ , BMA, and several values of  $P_0$  and  $n_0$ . The AUCs did not change with the introduction of a PP, regardless of the parameters considered.

### 4.2 Variable selection

For approach 1 that is used to generate simulated dose reductions, Figures 5 and 6, respectively, show the proportion of times where each variable is selected by the models and the proportion of times where each



**Figure 1.** Distribution of age, weight loss, bilirubin, ECOG performance status and treatment line for the  $n$  patients with complete observations over all cycles (sample I). (a) Age; (b) weight loss; (c) Bilirubin; (d) ECOG performance status; and (e) treatment line.

**Table 4.** Number of selected variables and AUCs for usual SSVS and SSVS using the power prior for the two approaches,  $a_0 \in \{0.1, 0.5, 1\}$ , BMA and several values of  $P_0$  and  $n_0$ .

		Number of selected variables			AUC		
					SSVS		
		5			0.69 (0.04)		
<b>PP – Approach 1</b>							
		$a_0 = 0.1$	$a_0 = 0.5$	$a_0 = 1$	$a_0 = 0.5$	$a_0 = 0.5$	$a_0 = 1$
$P_0 = P$	$n_0 = n_1$	3	6	9	0.68 (0.03)	0.68 (0.04)	0.68 (0.05)
$P_0 = P$	$n_0 = 1.77n_1$	3	6	10	0.68 (0.03)	0.68 (0.04)	0.68 (0.04)
$P_0 = 2P$	$n_0 = n_1$	2	8	13	0.69 (0.03)	0.68 (0.04)	0.68 (0.04)
$P_0 = 2P$	$n_0 = 1.77n_1$	2	10	14	0.69 (0.04)	0.69 (0.05)	0.69 (0.05)
$P_0 = P/2$	$n_0 = n_1$	3	6	11	0.69 (0.03)	0.69 (0.04)	0.68 (0.04)
$P_0 = P/2$	$n_0 = 1.77n_1$	2	4	9	0.68 (0.03)	0.68 (0.04)	0.68 (0.03)
<b>PP – Approach 2</b>							
		$a_0 = 0.1$	$a_0 = 0.5$	$a_0 = 1$	$a_0 = 0.1$	$a_0 = 0.5$	$a_0 = 1$
$P_0 = P$	$n_0 = n_1$	3	4	9	0.68 (0.03)	0.68 (0.04)	0.69 (0.05)
$P_0 = P$	$n_0 = 1.77n_1$	2	4	10	0.68 (0.03)	0.68 (0.05)	0.68 (0.05)
$P_0 = 2P$	$n_0 = n_1$	2	6	13	0.69 (0.03)	0.68 (0.04)	0.69 (0.05)
$P_0 = 2P$	$n_0 = 1.77n_1$	2	7	13	0.69 (0.03)	0.69 (0.05)	0.69 (0.05)
$P_0 = P/2$	$n_0 = n_1$	2	4	10	0.69 (0.03)	0.68 (0.04)	0.69 (0.04)
$P_0 = P/2$	$n_0 = 1.77n_1$	3	4	9	0.69 (0.03)	0.68 (0.03)	0.68 (0.03)

variable is selected by the models against values of mean weight elicited by clinicians. Figures 7 and 8 are the corresponding figures for approach 2. Tables 5 to 10 in Appendix 1 are the detailed results for all scenarios.

SSVS selects only a few variables. Only five variables were selected at least for five of the 10 cases conducted: bilirubin greater than  $35 \mu\text{mol/L}$  and greater than  $50 \mu\text{mol/L}$ , diarrhea of grade 2 and 3, and anemia of grade 2.

Variable selection varies according to the different settings for the power prior, more particularly with the power prior parameter. Variable selection did not change significantly with a varying sample size and dose reduction frequency of the simulated sample. When using the power prior approach, variables selected by the SSVS were also selected at least for five of the 10 experiences conducted, except anemia grade 2, for which the mean and maximal value for the elicited weights were, respectively, 12 and 50 with four out of the six physicians having chosen a null value for this variable.

For a small value of the power prior parameter, variables having high elicited weights were not selected in more than half of the cases. The number of selected variables with high elicited weights increased with the power prior parameter value. The power prior selects variables with high elicited weight and few observations, such as ECOG performance status grade 3 (11 observations) and vomiting grade 3 (2 observations). It also selects variables with high elicited weight but with a larger number of observations: ECOG performance status grade 4 (21 observations), asthenia grade 3 (21 observations) and age greater than 80 years (32 observations).

## 5 Discussion

Our method improves SSVS variable selection by automatically selecting relevant variables based on experts' everyday practice and discarding variables elicited by experts as not being important. Our method was also able to select rare variables that could not be selected using conventional data-driven approaches with relatively small samples. Of note AUC did not significantly vary across the different models and was moderate.

### 5.1 An automated and flexible variable selection approach

Once expertise is quantified by the elicited weights, our proposed method allows one to automatically select variables without the need for a consensus step. The two suggested approaches for generating simulated data based on experts' elicitation weights can provide different sets of selected variables, which may be explained by the randomness of the second approach in comparison to the first. Therefore, our method allows to incorporate elicitation variability in several ways and therefore can aggregate multiple expert opinions while taking into account their variability.

The number of variables included in the final model is flexible as it depends heavily on the numerical power prior parameter value, with 2 to 14 among 32 variables selected. Smaller values of the power prior parameter lead to fewer variables selected and discarding variables that were retrieved in the data-driven based analysis (SSVS). Larger values of this parameter lead to a larger number of selected variables combined both data and expert knowledge. Most of the SSVS selected variables were included, except anemia, for which the elicited weights were very low (mean weight: 33). Of note, bilirubin was selected as UGT1A1 polymorphism, closely linked to bilirubin level that is largely described as associated with Irinotecan adverse events.<sup>23</sup> The selected variables were strongly correlated with the elicited weights, with the variables having the highest elicited weight all included when using the largest power prior parameter. Consequently, we suggest repeating the procedure using several different values of the power prior parameter, such as  $a_0 \in \{1, 0.5, 0.1\}$ . On this basis,  $a_0$  can be chosen according to the number of relevant variables that one wishes to keep for building a decision tool, or guided by a trade-off between the selected variables stemming from the data and from the elicited expert knowledge.

### 5.2 An innovative approach to model medical decision

To our knowledge, power prior methods have been used mainly to vary the weight of historical data used in a current data analysis.<sup>17</sup> In this paper, we suggested to use a power prior to incorporate information in the form of clinical expertise by using simulated data obtained from elicited weights. This methodology appears promising and can be applied to a variety of situations involving medical decision making.

The proposed method aimed to model decision making. Personalized medicine requires the consolidation of knowledge about personalized medicine and the development of CDSS that provide actionable information.<sup>2</sup> Besides knowledge about disease subtypes and genomic variants, comprehensive understanding of the cognitive processes and parameters that influence medical decisions is needed. IT solutions must be proposed for data sharing and building a consensus on clinical interpretations of complex patient data. These solutions could integrate data reasoning capabilities together with computational data-driven

approaches like prediction tools, federated queries, and follow-up systems. Not surprisingly, our results show that the clinicians that participated in our study might act differently when dealing with the same clinical situation. Conventional knowledge-based CDSS and guidelines implementation in EHR systems are means to reduce variability in decision making. Our approach is an attempt to make explicit the parameters that influence clinical decisions about dose reduction in cancer treatment and explain data heterogeneity.

### 5.3 Elicitation step enables to capture practice variability

Predictive performance was moderate with an AUC of approximately 0.7 for all power prior parameters, which did not change with the addition of expert information. Two phenomena might explain this moderate performance: missing variables and practice variability among physicians. The 32 variables included in our study as potentially important for clinicians' decisions were those from the follow-up questionnaire used routinely to assign patients' doses. This assignment is performed the day before the patient venue for organizational purpose, based on this questionnaire. When the six clinicians who elicited their decision process in this study were asked about other variables not in the questionnaire that could influence doses, only two additional variables were mentioned: therapeutic objective (two clinicians) and intestinal obstruction (one clinician). Therefore, it is unlikely that important variables were not included in our study. By contrast, elicited weights deeply vary across physicians, in favor of practice variability among physicians. A study already concluded that large differences observed in practice probably result mostly from individual variation in clinical practice.<sup>3</sup> Therefore, the moderate performance achieved by our model is likely due to practice variability and this might hold for most use cases.

### 5.4 An approach limited by the contain of EHR

Our use case is based on monocentric data. It is indeed very difficult to gather real-world multicentric clinical data such as diarrhea, and vomiting grades due to the lack of interoperability of the different EHR across hospitals for clinical variables. Most multicentric real world data are based on administrative claims that do not contain all relevant information to model medical decision. As a perspective, it is of utmost importance to rely on interoperable EHR to have large sample size. Moreover, EHR contain is limited: it is noteworthy that medical decision process to reduce the dose is not described in EHR, that is, variables used for individual dose reduction are not reported.

### 5.5 The final perspective: dose regimen modeling to improve survival

Modeling dose regimen is an important intermediate step to model the complex relationship between patients and tumors characteristics, treatment and disease free survival. Indeed, the final aim of our approach is to assess which combination of dose will provide for each patient the best survival. Literature about the relevance of dose adjustment to improve survival is contradictory with some studies showing that reductions decrease survival while other shows no consequence on survival with reduced adverse events.<sup>8,24</sup> Finally, all of a patient's previous treatments and side effects may be considered by a physician to adapt the dose, not only those in the most recent cycle. Moreover, the time intervals between successive doses vary between patients, depending on their most recent toxicities, although time was not included in our analysis. Furthermore, past doses and administration schedules are major causes of toxicities, so there may be a complex causal effect, which has still to be modelled. While some real world studies highlighted strong difference in chemotherapy use, high frequency of dose adjustment and differential survival compared to clinical trials, no study tackled the complex relationship between dose adjustment and survival.<sup>25,26</sup> Few studies indeed focus on medical decision, and more particularly on dose regimen in routine data model building. This is, however, an essential step to assess drug efficacy. As a perspective, one should develop models that can take into account all these complex steps with the aim to find the best treatment for each patient.

## 6 Conclusion

In this article, we proposed a Bayesian method based on a power prior approach that allows one to combine information from patient data and elicited expert opinion when performing variable selection. In this method,

theoretical doses are generated based on the experts' elicited weights for variables, with a power prior to analyze the observed data and a simulated dataset based on the weights. Our application shows that this method enables one to select rare variables that cannot be selected using only patient data and to discard variables that appear to be relevant based on the data but not from the expert perspective. Combining expert opinion and data driven knowledge is a crucial step toward personalized medicine. Selection of the optimal set of important variables making in dose reduction decisions is essential to build decision-making tools that could help physicians to better manage their patients. Our method can be applied in all settings in which we want to account for experts' opinion when selecting variables.

### Authors' contributions

ASJ and SZ made equal contributions and are co-last authors.

### Acknowledgements

The authors thank Dr Anne-Laure Pointet and Dr Jean-Nicolas Vaillant from Georges Pompidou European Hospital for their help with elicitation. The simulations were performed at the HPCaVe at UPMC-Sorbonne Université.

### Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the French National Cancer Institut (INCa) [grant number 10801, 9539, SHSE SP 16-031].

### Ethics approval and consent to participate

The project was approved by the HEGP institutional review board CERHUPO (CPP Ile de France II, IRB registration #: 00001072) under the reference 2014-12-08.


### Availability of data and material

The database and its characteristics were supplied to the Data Protection Agency (CNIL). Data protection will follow the rules: the data will be confined at HEGP; control of data access will be performed through the HEGP infrastructure; deidentification algorithms will be used before data sharing.

### ORCID iD

Sandrine Boulet  <http://orcid.org/0000-0001-9863-5466>

Moreno Ursino  <http://orcid.org/0000-0002-5709-4322>

Peter Thall  <http://orcid.org/0000-0002-7293-529X>

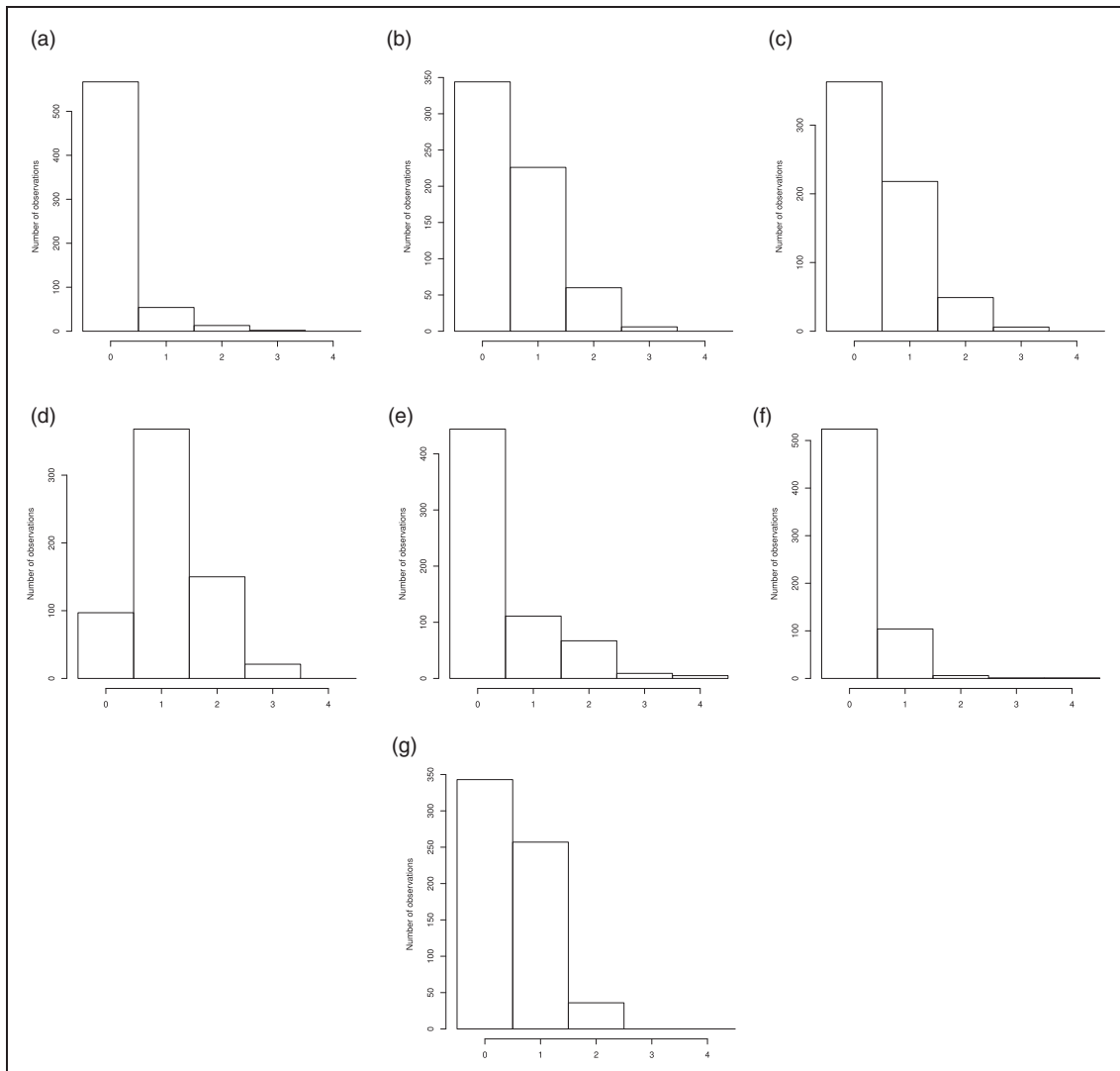
### References

1. Frankovich J, Longhurst CA and Sutherland SM. Evidence-based medicine in the EMR Era. *New Engl J Med* 2011; **365**: 1758–1759. <http://www.nejm.org/doi/abs/10.1056/NEJMp1108726>.
2. Tenenbaum JD, Avillach P, Benham-Hutchins M, et al. An informatics research agenda to support precision medicine: seven key areas. *J Am Med Informati Assoc* 2016; **23**: 791–795. <https://academic.oup.com/jamia/article/23/4/791/2198415>.
3. Smellie WSA. Methodology for constructing guidance. *J Clin Pathol* 2005; **58**: 249–253. <http://jcp.bmj.com/cgi/doi/10.1136/jcp.2004.018374>.

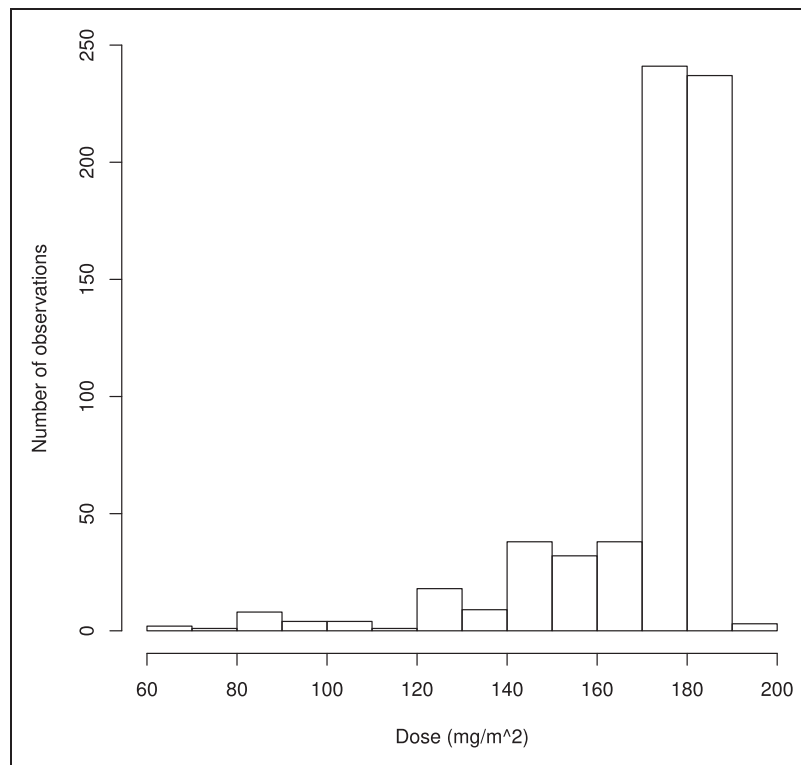
4. Zhuang ZY, Churilov L, Burstein F, et al. Combining data mining and case-based reasoning for intelligent decision support for pathology ordering by general practitioners. *Eur J Operation Res* 2009; **195**: 662–675. <http://linkinghub.elsevier.com/retrieve/pii/S0377221707010806>.
5. Caliskan A, Bryson JJ and Narayanan A. Semantics derived automatically from language corpora contain human-like biases. *Science* 2017; **356**: 183–186. <http://www.sciencemag.org/lookup/doi/10.1126/science.aal4230>.
6. Visa L, Jiménez-Fonseca P, Martínez EA, et al. Efficacy and safety of chemotherapy in older versus non-older patients with advanced gastric cancer: a real-world data, non-inferiority analysis. *J Geriatric Oncol* 2018; **9**: 254–264. <https://linkinghub.elsevier.com/retrieve/pii/S1879406817302473>.
7. Bhatnagar B, Gilmore S, Goloubeva O, et al. Chemotherapy dose reduction due to chemotherapy induced peripheral neuropathy in breast cancer patients receiving chemotherapy in the neoadjuvant or adjuvant settings: a single-center experience. *SpringerPlus* 2014; **3**: 366. <http://springerplus.springeropen.com/articles/10.1186/2193-1801-3-366>.
8. Stocker G, Hacker UT, Fiteni F, et al. Clinical consequences of chemotherapy dose reduction in obese patients with stage III colon cancer: a retrospective analysis from the PETACC 3 study. *Eur J Cancer* 2018; **99**: 49–57. <https://linkinghub.elsevier.com/retrieve/pii/S095980491830844X>.
9. Jansman FG, Sleijfer DT, Coenen JL, et al. Risk factors determining chemotherapeutic toxicity in patients with advanced colorectal cancer. *Drug Safety* 2000; **23**: 255–278. <http://link.springer.com/10.2165/00002018-200023040-00001>.
10. Tibshirani R. Regression shrinkage and selection via the lasso: a retrospective: regression shrinkage and selection via the Lasso. *J Royal Stat Soc: Ser B (Stat Methodol)* 2011; **73**: 273–282. <http://doi.wiley.com/10.1111/j.1467-9868.2011.00771.x>.
11. Fan J and Li R. Variable selection via nonconcave penalized likelihood and its oracle properties. *J Am Stat Assoc* 2001; **96**: 1348–1360. <http://www.jstor.org/stable/3085904>.
12. Efron B, Hastie T, Johnstone I, et al. Least angle regression. *Ann Stat* 2004; **32**: 407–451. <http://www.jstor.org/stable/3448465>.
13. Zou H and Hastie T. Regularization and variable selection via the elastic net. *J Royal Stat Soc: Ser B (Stat Methodol)* 2005; **67**: 301–320. <http://doi.wiley.com/10.1111/j.1467-9868.2005.00503.x>.
14. Bondell HD and Reich BJ. Simultaneous regression shrinkage, variable selection and clustering of predictors with OSCAR. *Biometrics* 2008; **64**: 115–123. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2605279/>.
15. Mitchell TJ and Beauchamp JJ. Bayesian variable selection in linear regression. *J Am Stat Assoc* 1988; **83**: 1023–1032. <http://www.jstor.org/stable/2290129>.
16. George EI and McCulloch RE. Variable selection via gibbs sampling. *J Am Stat Assoc* 1993; **88**: 881–889. <http://www.jstor.org/stable/2290777>.
17. Ibrahim JG and Chen MH. Power prior distributions for regression models. *Stat Sci* 2000; **15**: 46–60. <http://www.jstor.org/stable/2676676>.
18. Benchimol EI, Smeeth L, Guttman A, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. *PLOS Med* 2015; **12**: e1001885 <http://dx.plos.org/10.1371/journal.pmed.1001885>.
19. Zapletal E, Rodon N, Grabar N, et al. Methodology of integration of a clinical data warehouse with a clinical information system: the HEGP case. *Studies Health Technol Informat* 2010; **160**: 193–197.
20. Jannot AS, Zapletal E, Avillach P, et al. The Georges Pompidou University Hospital Clinical Data Warehouse: a 8-years follow-up experience. *Int J Med Informat* 2017; **102**: 21–28. <http://www.sciencedirect.com/science/article/pii/S1386505617300370>.
21. Bekele BN and Thall PF. Dose-finding based on multiple toxicities in a soft tissue sarcoma trial. *J Am Stat Assoc* 2004; **99**: 26–35. <http://www.jstor.org/stable/27590350>.
22. Raftery AE, Madigan D and Hoeting JA. Bayesian model averaging for linear regression models. *J Am Stat Assoc* 1997; **92**: 179–191. <http://www.jstor.org/stable/2291462>.
23. Innocenti F, Schilsky RL, Ramírez J, et al. Dose-finding and pharmacokinetic study to optimize the dosing of Irinotecan according to the *UGT1A1* genotype of patients with cancer. *J Clin Oncol* 2014; **32**: 2328–2334. <http://ascopubs.org/doi/10.1200/JCO.2014.55.2307>.
24. Yang JCH, Sequist LV, Zhou C, et al. Effect of dose adjustment on the safety and efficacy of afatinib for *EGFR* mutation-positive lung adenocarcinoma: *post hoc* analyses of the randomized LUX-Lung 3 and 6 trials. *Ann Oncol* 2016; **27**: 2103–2110. <https://academic.oup.com/annonc/article-lookup/doi/10.1093/annonc/mdw322>.
25. Sacher AG, Le LW, Lau A, et al. Real-world chemotherapy treatment patterns in metastatic non-small cell lung cancer: are patients undertreated?: Treatment Patterns in Metastatic NSCLC. *Cancer* 2015; **121**: 2562–2569.
26. Pericay C, Gallego J, Montes AF, et al. P-277real world data in colorectal cancer: a retrospective analysis of overall survival in metastatic colorectal cancer patients between 2011-2015 treated in Spain, preliminary results (RWD-ACROSS study). *Ann Oncol* 2018; **29**. <https://academic.oup.com/annonc/article/doi/10.1093/annonc/mdy151.276/5039537>.



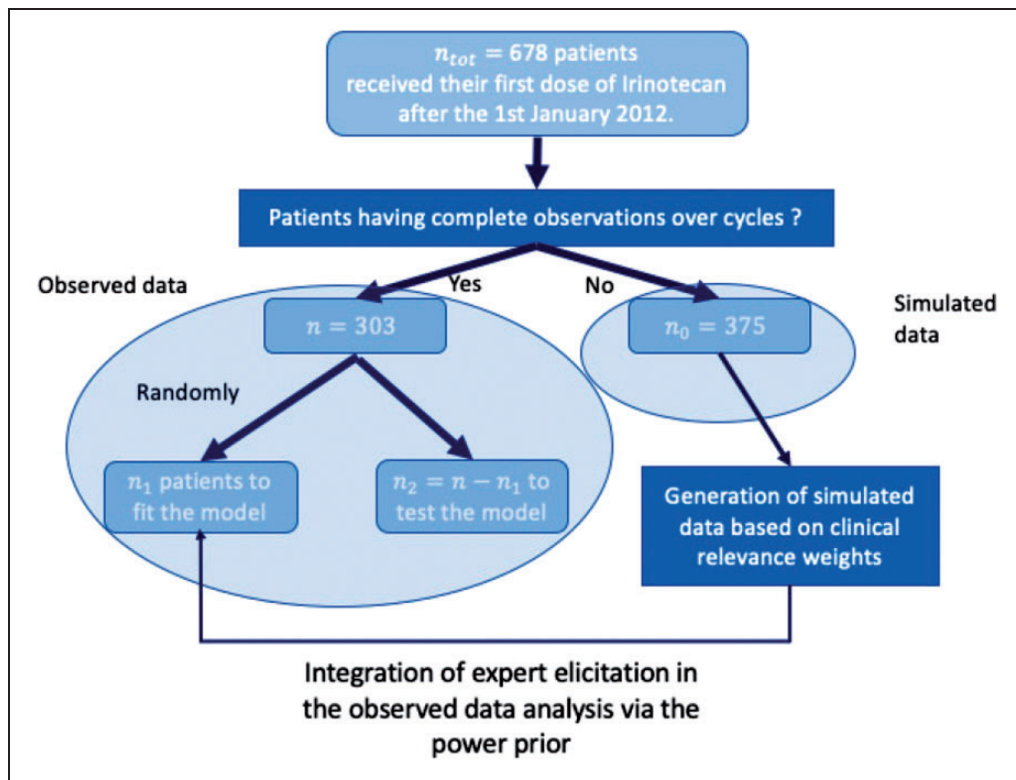
## Appendix I



**Figure 2.** Distribution of toxicities for the  $n$  patients with complete observations over all cycles. (a) Vomiting; (b) nausea; (c) diarrhea; (d) asthenia; (e) neutropenia; (f) thrombopenia; and (g) anemia.

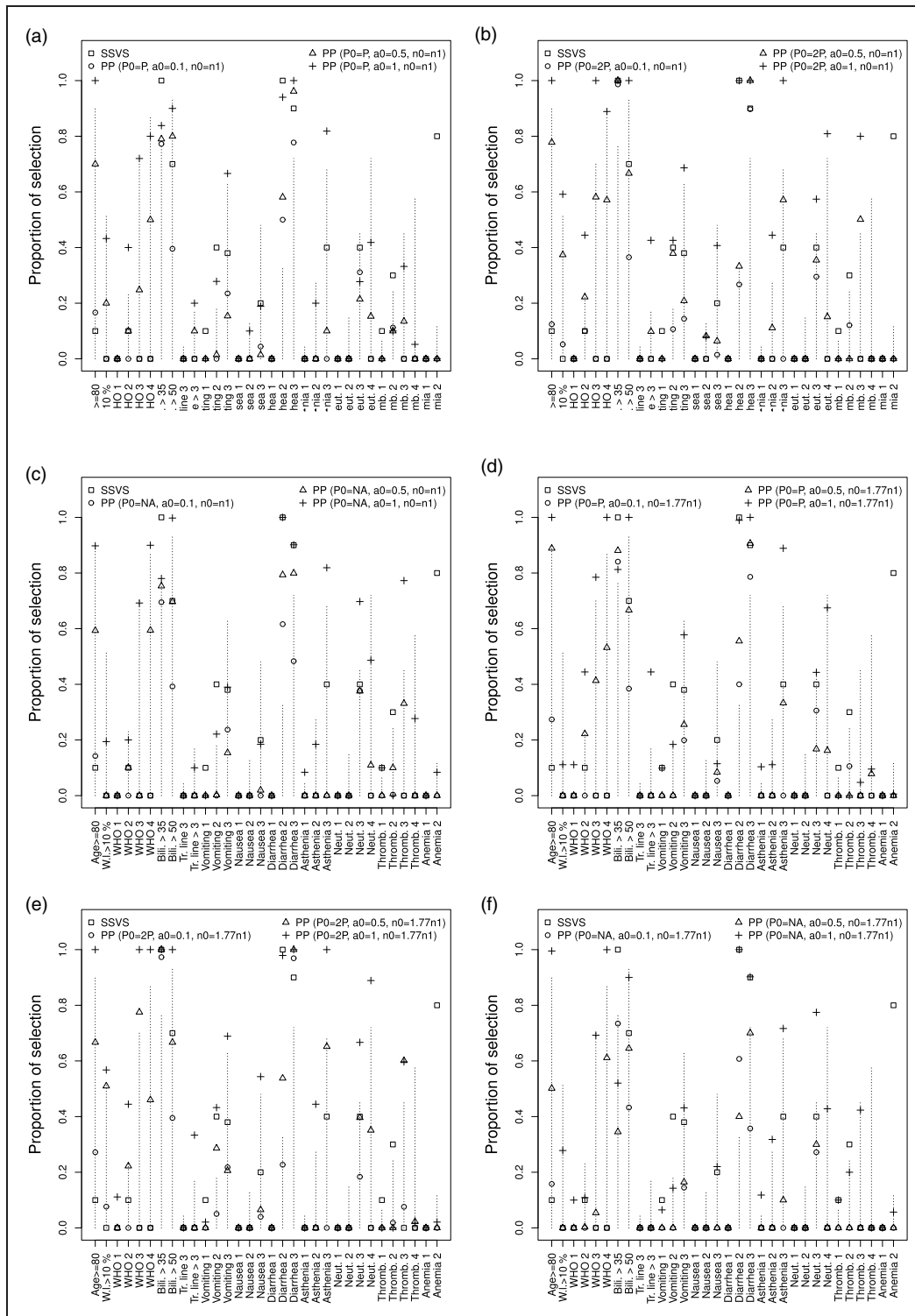


**Figure 3.** Distribution of doses for the  $n$  patients with complete observations over all cycles (sample 1).

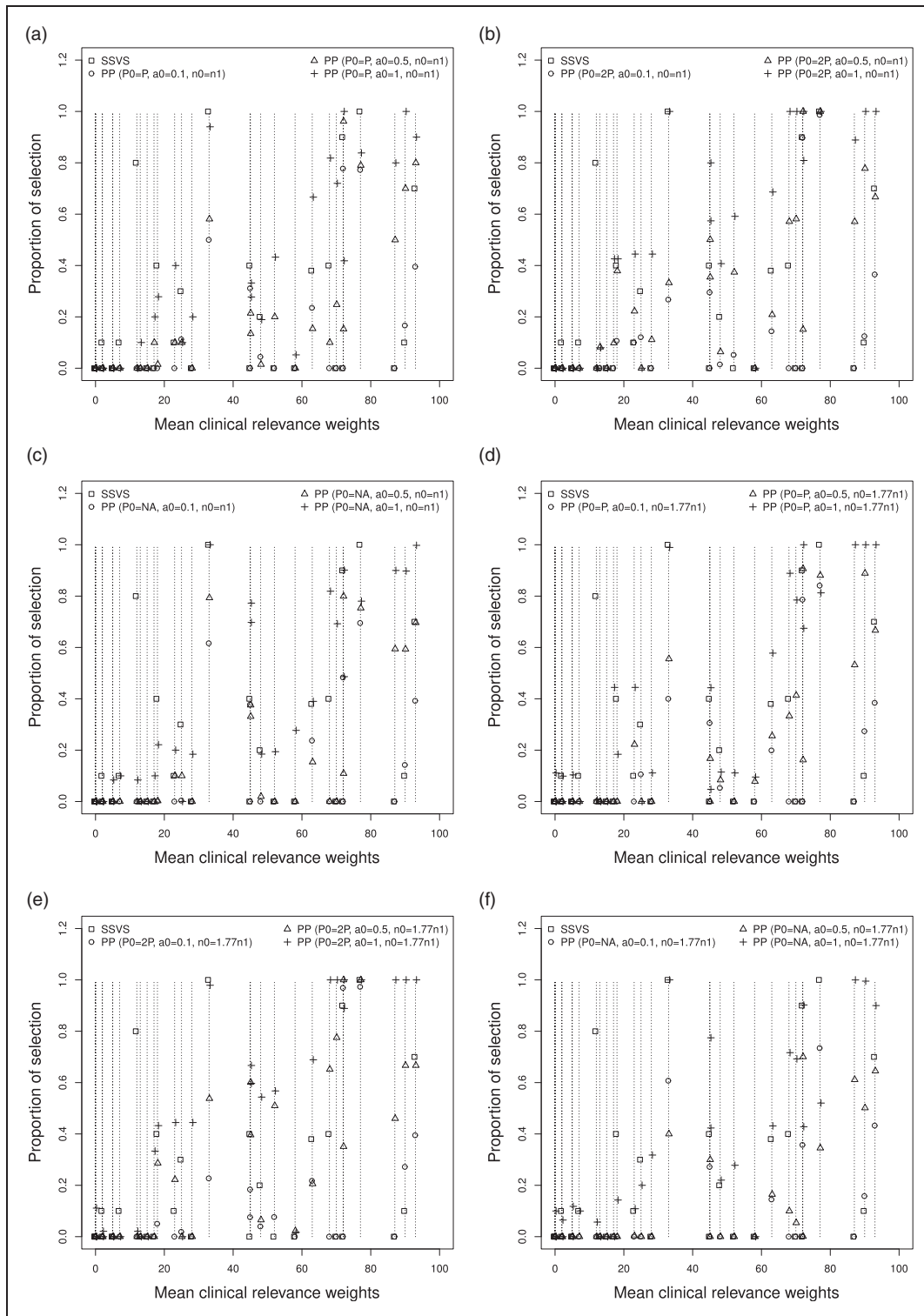


**Figure 4.** Flowchart of the use of patients.

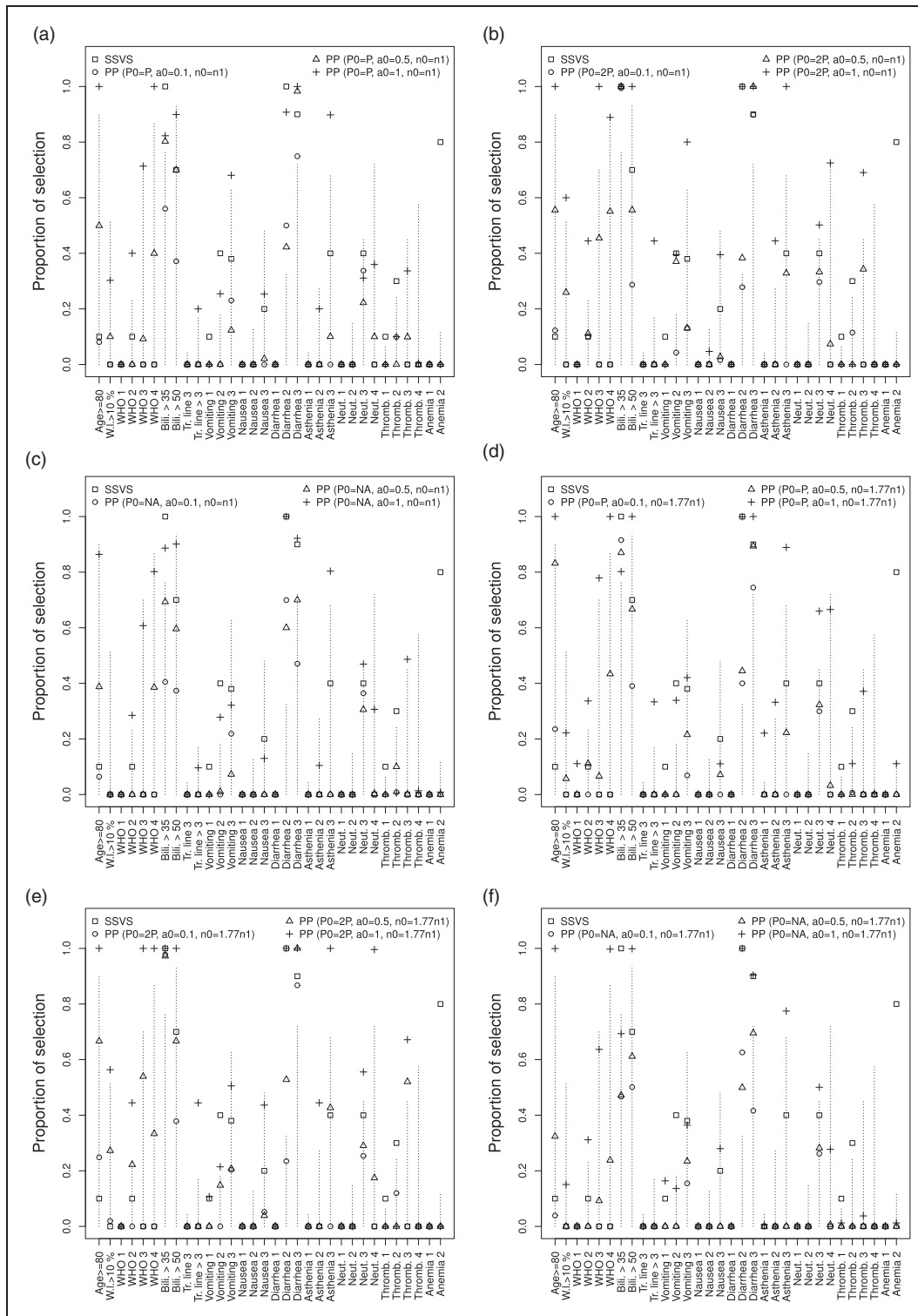




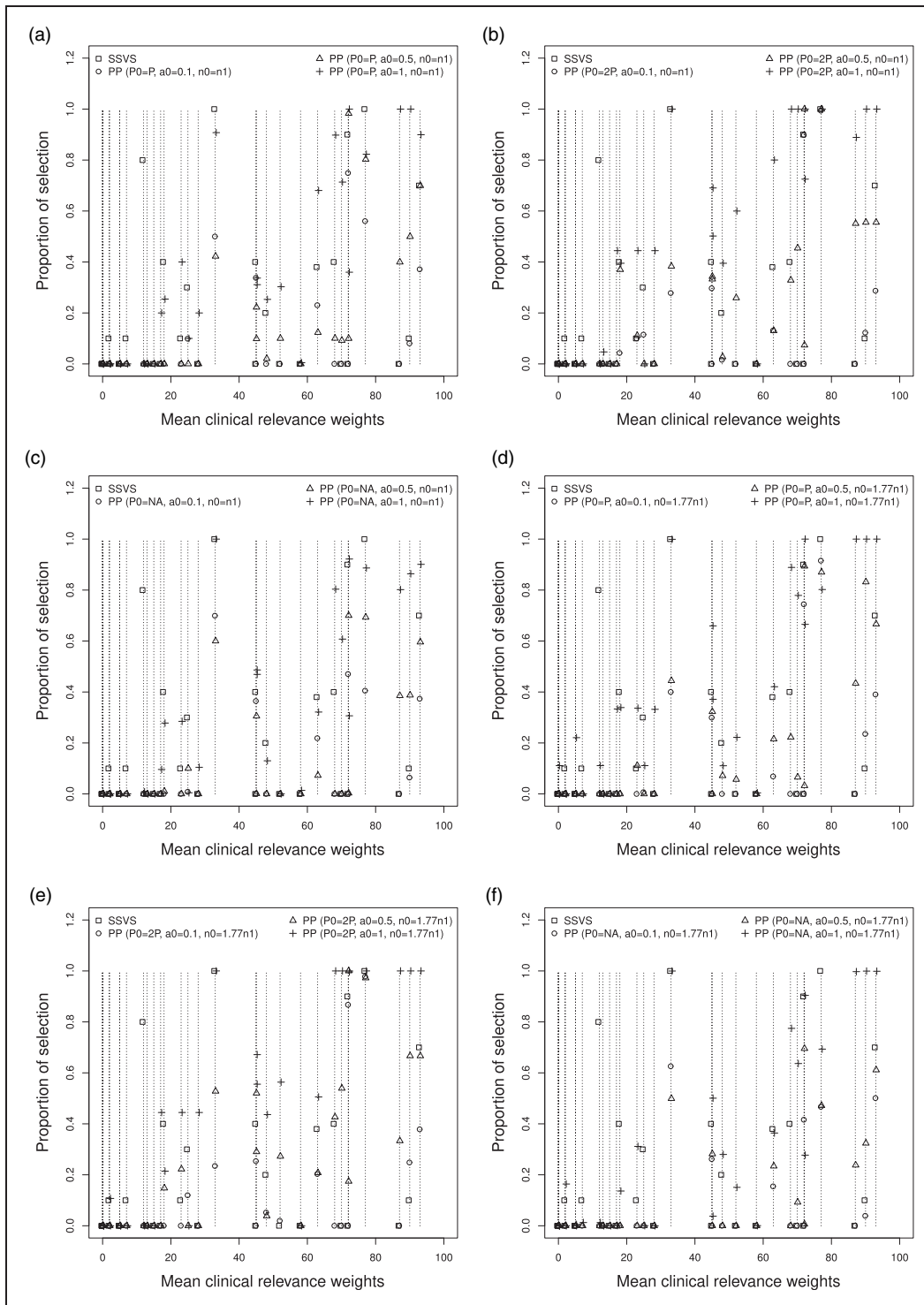
**Figure 5.** Proportion of times each variable is selected by the models for approach I, several values of  $P_0$ ,  $n_0$  and  $a_0$ . The values of mean weights elicited by clinicians for each variable are equal to the length of the dashed lines. Each model is labeled with a different symbol. (a)  $P_0 = P$ ,  $n_0 = n_1$ ; (b)  $P_0 = 2P$ ,  $n_0 = n_1$ ; (c)  $P_0 = P/2$ ,  $n_0 = n_1$ ; (d)  $P_0 = P$ ,  $n_0 = 1.77 \times n_1$ ; (e)  $P_0 = 2P$ ,  $n_0 = 1.77 \times n_1$ ; and (f)  $P_0 = P/2$ ,  $n_0 = 1.77 \times n_1$ .



**Figure 6.** Proportion of times each variable is selected by the models against values of mean weights elicited by clinicians for approach 1, several values of  $P_0$ ,  $n_0$  and  $a_0$ . Each model is labeled with a different symbol. (a)  $P_0 = P, n_0 = n_1$ ; (b)  $P_0 = 2P, n_0 = n_1$ ; (c)  $P_0 = P/2, n_0 = n_1$ ; (d)  $P_0 = P, n_0 = 1.77 \times n_1$ ; (e)  $P_0 = 2P, n_0 = 1.77 \times n_1$ ; and (f)  $P_0 = P/2, n_0 = 1.77 \times n_1$ .



**Figure 7.** Proportion of times each variable is selected by the models for approach 2, several values of  $P_0$ ,  $n_0$  and  $a_0$ . The values of mean weights elicited by clinicians for each variable are equal to the length of the dashed lines. Each model is labeled with a different symbol. (a)  $P_0 = P$ ,  $n_0 = n_1$ ; (b)  $P_0 = 2P$ ,  $n_0 = n_1$ ; (c)  $P_0 = P/2$ ,  $n_0 = n_1$ ; (d)  $P_0 = P$ ,  $n_0 = 1.77 \times n_1$ ; (e)  $P_0 = 2P$ ,  $n_0 = 1.77 \times n_1$ ; and (f)  $P_0 = P/2$ ,  $n_0 = 1.77 \times n_1$ .



**Figure 8.** Proportion of times each variable is selected by the models against values of mean weights elicited by clinicians for approach 2, several values of  $P_0$ ,  $n_0$  and  $a_0$ . Each model is labeled with a different symbol. (a)  $P_0 = P$ ,  $n_0 = n_1$ ; (b)  $P_0 = 2P$ ,  $n_0 = n_1$ ; (c)  $P_0 = P/2$ ,  $n_0 = n_1$ ; (d)  $P_0 = P$ ,  $n_0 = 1.77 \times n_1$ ; (e)  $P_0 = 2P$ ,  $n_0 = 1.77 \times n_1$ ; and (f)  $P_0 = P/2$ ,  $n_0 = 1.77 \times n_1$ .

**Table 5.** Estimated coefficients, number of selected variables and AUCs for usual SSVS and SSVS using the power prior for the first approach with  $P_0 = P$ ,  $a_0 \in \{0.1, 0.5, 1\}$ , BMA and several values of  $n_0$ .

Variables / Models	★	Power prior ( $P_0 = P$ )				$n_0 = n_1$		$n_0 = 1.77 \times n_1$	
		$a_0 = 0.1$	$a_0 = 0.5$	$a_0 = 0.1$	$a_0 = 0.5$	$a_0 = 0.1$	$a_0 = 0.5$	$a_0 = 0.1$	$a_0 = 0.5$
Intercept	–	–10.62 (1.36)	–6.33 (1.13)	–5.17 (0.45)	–6.33 (1.13)	–10.85 (1.55)	–5.05 (0.48)	–6.22 (0.93)	–12.55 (1.9)
Age $\geq 80$ years	5/32	1.63 <sup>0.1</sup> (0.68)	4.16 <sup>0.7</sup> (2.52)	1.59 <sup>0.2</sup> (0.79)	4.16 <sup>0.7</sup> (2.52)	7.95 <sup>1</sup> (3.66)	1.79 <sup>0.3</sup> (1.14)	4.59 <sup>0.9</sup> (2.51)	10.17 <sup>1</sup> (3.21)
Weight loss $> 10\%$	0/5	1.69 (0.64)	1.29 <sup>0.2</sup> (0.66)	0.89 (0.2)	1.29 <sup>0.2</sup> (0.66)	2.78 <sup>0.4</sup> (2.48)	0.84 (0.13)	0.98 (0.16)	1.86 <sup>0.1</sup> (0.74)
ECOG performance status 1	47/381	0.6 (0.23)	0.69 (0.34)	0.56 (0.2)	0.69 (0.34)	0.78 (0.29)	0.55 (0.18)	0.67 (0.27)	1.05 <sup>0.1</sup> (0.77)
ECOG performance status 2	25/100	1.87 <sup>0.1</sup> (0.48)	1.17 <sup>0.1</sup> (0.9)	1.08 (0.33)	1.17 <sup>0.1</sup> (0.9)	2.21 <sup>0.4</sup> (1.98)	1.13 (0.38)	1.32 <sup>0.2</sup> (0.81)	3.35 <sup>0.4</sup> (3.08)
ECOG performance status 3	2/11	1.61 (0.38)	2.39 <sup>0.2</sup> (1.87)	1.78 <sup>0.3</sup> (0.56)	2.39 <sup>0.2</sup> (1.87)	5.8 <sup>0.7</sup> (3.11)	0.99 (0.23)	2.86 <sup>0.4</sup> (1.6)	6.98 <sup>0.8</sup> (3.08)
ECOG performance status 4	3/21	1.24 (0.49)	2.9 <sup>0.5</sup> (1.83)	1.45 <sup>0.2</sup> (0.62)	2.9 <sup>0.5</sup> (1.83)	6.57 <sup>0.8</sup> (3.03)	1.06 (0.39)	2.99 <sup>0.5</sup> (1.64)	7.82 <sup>1</sup> (1.36)
Bilirubin $> 35 \mu\text{mol/L}$	1/1	5.94 <sup>1</sup> (0.38)	7.5 <sup>0.8</sup> (3.67)	5.38 <sup>0.8</sup> (0.79)	7.5 <sup>0.8</sup> (3.67)	10.77 <sup>0.8</sup> (4.04)	6.54 <sup>0.8</sup> (1.29)	9.29 <sup>0.9</sup> (2.59)	13.01 <sup>0.8</sup> (5.82)
Bilirubin $> 50 \mu\text{mol/L}$	2/5	4.39 <sup>0.7</sup> (1.64)	4.3 <sup>0.8</sup> (2.52)	2.37 <sup>0.4</sup> (1.17)	4.3 <sup>0.8</sup> (2.52)	8.11 <sup>0.9</sup> (3.59)	2.28 <sup>0.4</sup> (1.2)	4.26 <sup>0.7</sup> (2.59)	9.05 <sup>1</sup> (3.31)
Treatment line 3	8/46	1.28 (0.35)	0.55 (0.17)	0.61 (0.2)	0.55 (0.17)	0.7 (0.19)	0.61 (0.23)	0.55 (0.24)	0.76 (0.2)
Treatment line $> 3$	6/43	1.76 (0.6)	0.95 <sup>0.1</sup> (0.74)	0.77 (0.26)	0.95 <sup>0.1</sup> (0.74)	1.84 <sup>0.2</sup> (2.19)	0.77 (0.19)	1.08 (0.52)	3.25 <sup>0.4</sup> (2.84)
Vomiting 1	11/54	1.33 <sup>0.1</sup> (0.52)	0.78 (0.21)	0.7 (0.15)	0.78 (0.21)	1.03 (0.25)	0.69 (0.17)	0.83 (0.35)	1.31 <sup>0.1</sup> (0.66)
Vomiting 2	5/13	2.21 <sup>0.4</sup> (0.95)	0.96 (0.47)	0.91 (0.3)	0.96 (0.47)	1.43 <sup>0.3</sup> (0.78)	0.87 (0.29)	0.87 (0.5)	1.4 <sup>0.2</sup> (1.17)
Vomiting 3	1/2	3.69 <sup>0.4</sup> (1.86)	2.6 <sup>0.2</sup> (2.96)	2.08 <sup>0.2</sup> (1.81)	2.6 <sup>0.2</sup> (2.96)	4.72 <sup>0.7</sup> (3.1)	2.09 <sup>0.2</sup> (1.84)	2.49 <sup>0.3</sup> (2.54)	5.44 <sup>0.6</sup> (4.91)
Nausea 1	28/226	0.62 (0.13)	0.3 (0.11)	0.31 (0.07)	0.3 (0.11)	0.4 (0.15)	0.29 (0.07)	0.29 (0.12)	0.45 (0.13)
Nausea 2	13/60	0.63 (0.12)	0.5 (0.09)	0.47 (0.07)	0.5 (0.09)	0.73 <sup>0.1</sup> (0.53)	0.45 (0.08)	0.44 (0.13)	0.53 (0.15)
Nausea 3	2/6	2.21 <sup>0.2</sup> (2.27)	1.08 (0.45)	0.95 (0.51)	1.08 (0.45)	2.28 <sup>0.2</sup> (2.34)	0.94 <sup>0.1</sup> (0.52)	1.14 <sup>0.1</sup> (1)	2.13 <sup>0.1</sup> (3)
Diarrhea 1	28/218	0.48 (0.11)	0.29 (0.1)	0.29 (0.1)	0.29 (0.1)	0.37 (0.14)	0.28 (0.09)	0.26 (0.11)	0.33 (0.09)
Diarrhea 2	18/49	6.54 <sup>1</sup> (1.88)	2.47 <sup>0.6</sup> (0.99)	2.16 <sup>0.5</sup> (0.61)	2.47 <sup>0.6</sup> (0.99)	4.31 <sup>0.9</sup> (1.21)	2.03 <sup>0.4</sup> (0.6)	2.31 <sup>0.6</sup> (0.78)	5.2 <sup>1</sup> (1.27)
Diarrhea 3	4/6	6.49 <sup>0.9</sup> (2.48)	5.89 <sup>1</sup> (2.76)	3.64 <sup>0.8</sup> (1.22)	5.89 <sup>1</sup> (2.76)	9.46 <sup>1</sup> (2.97)	3.42 <sup>0.8</sup> (1.23)	5.7 <sup>0.9</sup> (2.47)	10.84 <sup>1</sup> (4.28)
Asthenia 1	41/368	0.66 (0.17)	0.62 (0.17)	0.5 (0.11)	0.62 (0.17)	0.88 (0.23)	0.51 (0.12)	0.67 (0.23)	1.09 <sup>0.1</sup> (0.64)
Asthenia 2	32/150	0.83 (0.18)	0.87 (0.36)	0.7 (0.21)	0.87 (0.36)	1.65 <sup>0.2</sup> (1.15)	0.72 (0.25)	0.86 (0.49)	1.78 <sup>0.1</sup> (1.85)
Asthenia 3	4/21	2.26 <sup>0.4</sup> (0.8)	1.55 <sup>0.1</sup> (0.53)	1.02 (0.2)	1.55 <sup>0.1</sup> (0.53)	3.92 <sup>0.8</sup> (1.75)	1.04 (0.2)	1.94 <sup>0.3</sup> (0.77)	5.98 <sup>0.9</sup> (2.77)
Neutropenia 1	12/111	0.65 (0.15)	0.4 (0.14)	0.39 (0.09)	0.4 (0.14)	0.51 (0.18)	0.37 (0.08)	0.38 (0.11)	0.54 (0.19)
Neutropenia 2	9/67	0.65 (0.05)	0.48 (0.1)	0.46 (0.07)	0.48 (0.1)	0.63 (0.14)	0.44 (0.07)	0.41 (0.08)	0.54 (0.11)
Neutropenia 3	4/9	3.73 <sup>0.4</sup> (2.12)	1.81 <sup>0.2</sup> (1.16)	1.68 <sup>0.3</sup> (0.83)	1.81 <sup>0.2</sup> (1.16)	2.93 <sup>0.3</sup> (2.89)	1.69 <sup>0.3</sup> (0.85)	1.55 <sup>0.2</sup> (0.71)	2.71 <sup>0.4</sup> (1.48)
Neutropenia 4	1/5	1.81 (0.48)	1.55 <sup>0.2</sup> (1.09)	1.05 (0.24)	1.55 <sup>0.2</sup> (1.09)	3.23 <sup>0.4</sup> (1.85)	1.09 (0.23)	1.85 <sup>0.2</sup> (1.25)	4.71 <sup>0.7</sup> (2.8)
Thrombopenia 1	15/104	1.49 <sup>0.1</sup> (0.61)	0.52 (0.13)	0.52 (0.13)	0.52 (0.13)	0.96 (0.35)	0.49 (0.1)	0.5 (0.13)	1.07 (0.37)
Thrombopenia 2	2/6	3.22 <sup>0.3</sup> (1.35)	0.71 <sup>0.1</sup> (0.53)	1.05 <sup>0.1</sup> (0.7)	0.71 <sup>0.1</sup> (0.53)	1.1 <sup>0.1</sup> (1.6)	0.96 <sup>0.1</sup> (0.61)	0.48 (0.12)	0.57 (0.14)
Thrombopenia 3	0/1	3.21 (0.52)	1.55 <sup>0.1</sup> (0.67)	1.36 (0.17)	1.55 <sup>0.1</sup> (0.67)	2.99 <sup>0.3</sup> (2.73)	1.29 (0.17)	1.17 (0.27)	1.81 (0.61)
Thrombopenia 4	0/1	1.13 (0.11)	0.9 (0.11)	0.89 (0.04)	0.9 (0.11)	1.03 <sup>0.1</sup> (0.31)	0.91 (0.05)	1.09 <sup>0.1</sup> (0.4)	1.63 <sup>0.1</sup> (1.21)
Anemia 1	26/257	0.42 (0.09)	0.19 (0.04)	0.19 (0.04)	0.19 (0.05)	0.26 (0.07)	0.18 (0.03)	0.18 (0.03)	0.3 (0.07)
Anemia 2	12/36	3.25 <sup>0.8</sup> (1.09)	0.9 (0.19)	1.18 (0.26)	0.9 (0.19)	1.08 (0.35)	1.16 (0.24)	0.86 (0.23)	1.03 (0.32)
Number of selected variables	–	5	6	3	6	9	2	6	10
AUC	–	0.69 (0.04)	0.68 (0.03)	0.68 (0.03)	0.68 (0.04)	0.68 (0.05)	0.68 (0.03)	0.68 (0.04)	0.68 (0.04)

Note: <sup>s</sup>,  $0 \leq s \leq 1$  is the proportion of times a variable is selected by the model. The \* column indicates the proportion of dose reductions linked to the variable.

**Table 6.** Estimated coefficients, number of selected variables and AUCs for usual SSVS and SSVS using the power prior for the first approach with  $P_0 = 2P$ ,  $a_0 \in \{0.1, 0.5, 1\}$ , BMA and several values of  $n_0$ .

Variables / Models	★	Power prior ( $P_0 = 2P$ )						$n_0 = n_1$ $a_0 = 1$	$n_0 = 1.77 \times n_1$ $a_0 = 1$		
		SSVS		$a_0 = 0.5$		$a_0 = 0.1$				$a_0 = 0.5$	
		$a_0 = 0.1$	$a_0 = 0.5$	$a_0 = 0.1$	$a_0 = 0.5$	$a_0 = 0.1$	$a_0 = 0.5$			$a_0 = 0.1$	$a_0 = 0.5$
Intercept	-	-10.62 (1.36)	-5.64 (0.78)	-11.45 (2.7)	-4.43 (0.39)	-5.61 (0.93)	-11.45 (2.7)	-4.43 (0.39)	-5.61 (0.93)		
Age $\geq 80$ years	5/32	1.63 <sup>0.1</sup> (0.68)	4.05 <sup>0.8</sup> (1.9)	9.71 (2.66)	1.43 <sup>0.3</sup> (0.74)	4.27 <sup>0.7</sup> (2.17)	9.71 (2.66)	1.43 <sup>0.3</sup> (0.74)	4.27 <sup>0.7</sup> (2.17)		
Weight loss $> 10\%$	0/5	1.69 (0.64)	2.35 <sup>0.4</sup> (1.35)	4.93 <sup>0.6</sup> (2.83)	1.26 <sup>0.1</sup> (0.56)	2.91 <sup>0.5</sup> (2.22)	4.93 <sup>0.6</sup> (2.83)	1.26 <sup>0.1</sup> (0.56)	2.91 <sup>0.5</sup> (2.22)		
ECOG performance status 1	47/381	0.6 (0.23)	0.54 (0.24)	0.64 (0.24)	0.5 (0.19)	0.54 (0.24)	0.64 (0.24)	0.5 (0.19)	0.54 (0.24)		
ECOG performance status 2	25/100	1.87 <sup>0.1</sup> (0.48)	1.39 <sup>0.2</sup> (1.13)	3.11 <sup>0.4</sup> (3.41)	1.27 <sup>0.1</sup> (0.47)	1.39 <sup>0.2</sup> (1.13)	3.11 <sup>0.4</sup> (3.41)	1.28 (0.48)	1.38 <sup>0.2</sup> (0.97)		
ECOG performance status 3	2/11	1.61 (0.38)	3.15 <sup>0.6</sup> (1.56)	8.55 <sup>1</sup> (2.59)	1.07 (0.33)	3.52 <sup>0.8</sup> (1.78)	8.55 <sup>1</sup> (2.59)	1.11 (0.23)	3.52 <sup>0.8</sup> (1.78)		
ECOG performance status 4	3/21	1.24 (0.49)	2.47 <sup>0.6</sup> (1.14)	6.54 <sup>0.9</sup> (2.61)	0.96 (0.19)	2.39 <sup>0.5</sup> (1.09)	6.54 <sup>0.9</sup> (2.61)	0.98 (0.22)	2.39 <sup>0.5</sup> (1.09)		
Bilirubin $> 35 \mu\text{mol/L}$	1/1	5.94 <sup>1</sup> (0.38)	9.39 <sup>1</sup> (2.92)	13.04 <sup>1</sup> (2.56)	6.82 <sup>1</sup> (0.83)	9.65 <sup>1</sup> (2.88)	13.04 <sup>1</sup> (2.56)	6.49 <sup>1</sup> (0.84)	9.65 <sup>1</sup> (2.88)		
Bilirubin $> 50 \mu\text{mol/L}$	2/5	4.39 <sup>0.7</sup> (1.64)	4.13 <sup>0.7</sup> (2.33)	9.58 <sup>1</sup> (3.22)	2.15 <sup>0.4</sup> (0.99)	4.36 <sup>0.7</sup> (2.51)	9.58 <sup>1</sup> (3.22)	2 <sup>0.4</sup> (0.95)	4.36 <sup>0.7</sup> (2.51)		
Treatment line 3	8/46	1.28 (0.35)	0.52 (0.2)	0.62 (0.16)	0.56 (0.19)	0.52 (0.2)	0.62 (0.16)	0.51 (0.21)	0.47 (0.22)		
Treatment line $> 3$	6/43	1.76 (0.6)	1.08 <sup>0.1</sup> (0.65)	3.18 <sup>0.4</sup> (3.08)	0.71 (0.18)	1.08 <sup>0.1</sup> (0.65)	3.18 <sup>0.4</sup> (3.08)	0.65 (0.13)	0.8 (0.37)		
Vomiting 1	11/54	1.33 <sup>0.1</sup> (0.52)	0.71 (0.25)	1.05 (0.31)	0.62 (0.16)	0.71 (0.25)	1.05 (0.31)	0.6 (0.17)	0.7 (0.33)		
Vomiting 2	5/13	2.21 <sup>0.4</sup> (0.95)	2.3 <sup>0.4</sup> (2)	4.53 <sup>0.4</sup> (4.42)	1.22 <sup>0.1</sup> (0.67)	2.3 <sup>0.4</sup> (2)	4.53 <sup>0.4</sup> (4.42)	1.03 <sup>0.1</sup> (0.37)	1.54 <sup>0.3</sup> (1.46)		
Vomiting 3	1/2	3.69 <sup>0.4</sup> (1.86)	2.98 <sup>0.2</sup> (2.5)	6.42 <sup>0.7</sup> (3.33)	2.03 <sup>0.1</sup> (1.62)	2.98 <sup>0.2</sup> (2.5)	6.42 <sup>0.7</sup> (3.33)	2.1 <sup>0.2</sup> (1.72)	2.62 <sup>0.2</sup> (2.26)		
Nausea 1	28/226	0.62 (0.13)	0.28 (0.1)	0.39 (0.11)	0.29 (0.06)	0.28 (0.1)	0.39 (0.11)	0.26 (0.07)	0.27 (0.13)		
Nausea 2	13/60	0.63 (0.12)	0.53 <sup>0.1</sup> (0.44)	1.08 <sup>0.1</sup> (1.44)	0.42 (0.06)	0.53 <sup>0.1</sup> (0.44)	1.08 <sup>0.1</sup> (1.44)	0.39 (0.06)	0.34 (0.1)		
Nausea 3	2/6	2.21 <sup>0.2</sup> (2.27)	1.27 <sup>0.1</sup> (1.09)	2.98 <sup>0.4</sup> (2.5)	0.89 (0.41)	1.27 <sup>0.1</sup> (1.09)	2.98 <sup>0.4</sup> (2.5)	0.9 (0.39)	1.36 <sup>0.1</sup> (1.29)		
Diarrhea 1	28/218	0.48 (0.11)	0.32 (0.13)	0.41 (0.13)	0.31 (0.09)	0.32 (0.13)	0.41 (0.13)	0.29 (0.09)	0.29 (0.13)		
Diarrhea 2	18/49	6.54 <sup>1</sup> (1.88)	2.35 <sup>0.3</sup> (1.31)	5.45 <sup>1</sup> (2.16)	1.79 <sup>0.2</sup> (0.56)	2.35 <sup>0.3</sup> (1.31)	5.45 <sup>1</sup> (2.16)	1.79 <sup>0.2</sup> (0.48)	2.82 <sup>0.5</sup> (1.57)		
Diarrhea 3	4/6	6.49 <sup>0.9</sup> (2.48)	7.67 <sup>1</sup> (2.51)	15.08 <sup>1</sup> (3.69)	4.31 <sup>0.9</sup> (1.37)	7.67 <sup>1</sup> (2.51)	15.08 <sup>1</sup> (3.69)	4.28 <sup>1</sup> (1.43)	7.83 <sup>1</sup> (3.14)		
Asthenia 1	41/368	0.66 (0.17)	0.49 (0.15)	0.84 (0.39)	0.39 (0.07)	0.49 (0.15)	0.84 (0.39)	0.4 (0.1)	0.53 (0.21)		
Asthenia 2	32/150	0.83 (0.18)	1.17 <sup>0.1</sup> (0.79)	3.1 <sup>0.4</sup> (2.81)	0.66 (0.23)	1.17 <sup>0.1</sup> (0.79)	3.1 <sup>0.4</sup> (2.81)	0.63 (0.21)	1.12 (0.69)		
Asthenia 3	4/21	2.26 <sup>0.4</sup> (0.8)	2.57 <sup>0.6</sup> (1.27)	7.64 <sup>1</sup> (2.02)	0.98 (0.21)	2.57 <sup>0.6</sup> (1.27)	7.64 <sup>1</sup> (2.02)	1.05 (0.29)	2.96 <sup>0.7</sup> (1.48)		
Neutropenia 1	12/111	0.65 (0.15)	0.35 (0.09)	0.55 (0.15)	0.34 (0.07)	0.35 (0.09)	0.55 (0.15)	0.33 (0.05)	0.3 (0.04)		
Neutropenia 2	9/67	0.65 (0.05)	0.41 (0.09)	0.52 (0.13)	0.42 (0.07)	0.41 (0.09)	0.52 (0.13)	0.39 (0.07)	0.33 (0.05)		
Neutropenia 3	4/9	3.73 <sup>0.4</sup> (2.12)	2.58 <sup>0.4</sup> (1.79)	4.92 <sup>0.6</sup> (3.64)	1.77 <sup>0.3</sup> (0.96)	2.58 <sup>0.4</sup> (1.79)	4.92 <sup>0.6</sup> (3.64)	1.53 <sup>0.2</sup> (0.69)	2.98 <sup>0.4</sup> (2.13)		
Neutropenia 4	1/5	1.81 (0.48)	1.95 <sup>0.2</sup> (1.06)	5.94 <sup>0.8</sup> (2.54)	1.05 (0.26)	1.95 <sup>0.2</sup> (1.06)	5.94 <sup>0.8</sup> (2.54)	1.13 (0.34)	2.93 <sup>0.4</sup> (2.22)		
Thrombopenia 1	15/104	1.49 <sup>0.1</sup> (0.61)	0.4 (0.08)	0.78 (0.21)	0.41 (0.08)	0.4 (0.08)	0.78 (0.21)	0.38 (0.09)	0.38 (0.11)		
Thrombopenia 2	2/6	3.22 <sup>0.3</sup> (1.35)	0.62 (0.17)	0.69 (0.24)	1.38 <sup>0.1</sup> (1.26)	0.62 (0.17)	0.69 (0.24)	0.89 (0.17)	0.52 (0.11)		
Thrombopenia 3	0/1	3.21 (0.52)	4.04 <sup>0.5</sup> (3.02)	8.51 <sup>0.8</sup> (5.67)	1.51 (0.26)	4.04 <sup>0.5</sup> (3.02)	8.51 <sup>0.8</sup> (5.67)	1.48 <sup>0.1</sup> (0.29)	3.82 <sup>0.6</sup> (2.57)		
Thrombopenia 4	0/1	1.13 (0.11)	1 (0.25)	1.39 (0.55)	0.91 (0.06)	1 (0.25)	1.39 (0.55)	0.93 (0.06)	1.07 (0.19)		
Anemia 1	26/257	0.42 (0.09)	0.17 (0.05)	0.28 (0.09)	0.17 (0.03)	0.17 (0.05)	0.28 (0.09)	0.16 (0.02)	0.16 (0.03)		
Anemia 2	12/36	3.25 <sup>0.8</sup> (1.09)	0.83 (0.22)	1.02 (0.29)	1.12 (0.27)	0.83 (0.22)	1.02 (0.29)	1.04 (0.21)	0.82 (0.22)		
Number of selected variables	-	5	8	13	2	8	13	2	10		
AUC	-	0.69 (0.04)	0.68 (0.04)	0.68 (0.04)	0.69 (0.03)	0.68 (0.04)	0.68 (0.04)	0.69 (0.04)	0.69 (0.05)		

Note: <sup>s</sup>,  $0 \leq s \leq 1$  is the proportion of times a variable is selected by the model. The ★ column indicates the proportion of dose reductions linked to the variable.

**Table 7.** Estimated coefficients, number of selected variables and AUCs for usual SSVS and SSVS using the power prior for the first approach with  $P_0 = P/2$ ,  $\alpha_0 \in \{0.1, 0.5, 1\}$ , BMA and several values of  $n_0$ .

Variables / Models	*	Power prior ( $P_0 = P/2$ )				$n_0 = n_1$		$n_0 = 1.77 \times n_1$	
		SSVS	$\alpha_0 = 0.1$	$\alpha_0 = 0.5$	$\alpha_0 = 1$	$\alpha_0 = 0.1$	$\alpha_0 = 0.5$	$\alpha_0 = 0.1$	$\alpha_0 = 0.5$
Intercept	-	-10.62 (1.36)	-5.39 (0.41)	-6.54 (0.9)	-12.14 (2)	-5.32 (0.42)	-6.16 (1.3)	-12.11 (1.77)	
Age $\geq 80$ years	5/32	1.63 <sup>0.1</sup> (0.68)	1.45 <sup>0.1</sup> (0.66)	4.16 <sup>0.6</sup> (2.91)	7.74 <sup>0.9</sup> (4.51)	1.54 <sup>0.2</sup> (0.75)	4.14 <sup>0.5</sup> (3.46)	8.92 <sup>1</sup> (5.08)	
Weight loss $> 10\%$	0/5	1.69 (0.64)	0.82 (0.16)	0.9 (0.27)	2.03 <sup>0.2</sup> (1.29)	0.86 (0.19)	1.11 (0.42)	2.16 <sup>0.3</sup> (1.49)	
ECOG performance status 1	47/381	0.6 (0.23)	0.56 (0.18)	0.7 (0.25)	0.84 (0.38)	0.54 (0.17)	0.57 (0.22)	0.82 <sup>0.1</sup> (0.68)	
ECOG performance status 2	25/100	1.87 <sup>0.1</sup> (0.48)	1 (0.27)	0.96 <sup>0.1</sup> (0.69)	1.77 <sup>0.2</sup> (2.35)	0.92 (0.2)	0.57 (0.17)	1.18 <sup>0.1</sup> (1.44)	
ECOG performance status 3	2/11	1.61 (0.38)	1.78 <sup>0.3</sup> (0.56)	1.2 (0.42)	3.56 <sup>0.7</sup> (1.21)	0.88 (0.12)	1.15 <sup>0.1</sup> (0.39)	3.49 <sup>0.7</sup> (1.85)	
ECOG performance status 4	3/21	1.24 (0.49)	1.45 <sup>0.2</sup> (0.62)	3.28 <sup>0.6</sup> (1.93)	8.05 <sup>0.9</sup> (3.84)	1.07 (0.27)	3.1 <sup>0.6</sup> (1.64)	8.08 <sup>1</sup> (2.23)	
Bilirubin $> 35 \mu\text{mol/L}$	1/1	5.94 <sup>1</sup> (0.38)	4.79 <sup>0.7</sup> (1.79)	7.22 <sup>0.8</sup> (4.72)	9.51 <sup>1</sup> (5.4)	5.91 <sup>0.7</sup> (0.99)	5.06 <sup>0.3</sup> (5.6)	9.15 <sup>0.5</sup> (7.34)	
Bilirubin $> 50 \mu\text{mol/L}$	2/5	4.39 <sup>0.7</sup> (1.64)	2.41 <sup>0.4</sup> (1.04)	4.54 <sup>0.7</sup> (2.59)	9.55 <sup>1</sup> (4.15)	2.35 <sup>0.4</sup> (0.97)	4.33 <sup>0.6</sup> (2.84)	8.76 <sup>0.9</sup> (4.53)	
Treatment line 3	8/46	1.28 (0.35)	0.65 (0.2)	0.61 (0.2)	0.77 (0.19)	0.64 (0.21)	0.58 (0.29)	0.79 (0.34)	
Treatment line $> 3$	6/43	1.76 (0.6)	0.72 (0.16)	0.75 (0.24)	1.12 <sup>0.1</sup> (0.63)	0.68 (0.14)	0.62 (0.15)	0.87 (0.15)	
Vomiting 1	11/54	1.33 <sup>0.1</sup> (0.52)	0.73 (0.15)	0.83 (0.23)	1.25 (0.44)	0.74 (0.16)	0.87 (0.21)	1.55 <sup>0.1</sup> (0.39)	
Vomiting 2	5/13	2.21 <sup>0.4</sup> (0.95)	0.9 (0.26)	0.89 (0.41)	1.78 <sup>0.2</sup> (1.56)	0.84 (0.24)	0.74 (0.31)	1.19 <sup>0.1</sup> (1.02)	
Vomiting 3	1/2	3.69 <sup>0.4</sup> (1.86)	1.99 <sup>0.2</sup> (1.47)	1.95 <sup>0.2</sup> (1.48)	4.21 <sup>0.4</sup> (3.35)	1.93 <sup>0.1</sup> (1.33)	1.97 <sup>0.2</sup> (1.38)	4.49 <sup>0.4</sup> (3.15)	
Nausea 1	28/226	0.62 (0.13)	0.32 (0.06)	0.31 (0.09)	0.46 (0.11)	0.3 (0.06)	0.29 (0.1)	0.44 (0.12)	
Nausea 2	13/60	0.63 (0.12)	0.48 (0.07)	0.5 (0.06)	0.71 (0.25)	0.48 (0.07)	0.52 (0.05)	0.7 (0.17)	
Nausea 3	2/6	2.21 <sup>0.2</sup> (2.27)	0.93 (0.4)	0.97 (0.31)	1.93 <sup>0.2</sup> (2.02)	0.89 (0.38)	0.92 (0.3)	1.7 <sup>0.2</sup> (1.32)	
Diarrhea 1	28/218	0.48 (0.11)	0.29 (0.09)	0.27 (0.12)	0.34 (0.1)	0.28 (0.08)	0.26 (0.1)	0.34 (0.09)	
Diarrhea 2	18/49	6.54 <sup>1</sup> (1.88)	2.32 <sup>0.6</sup> (0.58)	2.78 <sup>0.8</sup> (0.91)	5.61 <sup>1</sup> (1.14)	2.26 <sup>0.6</sup> (0.59)	2.5 <sup>0.4</sup> (1.09)	5.44 <sup>1</sup> (1.2)	
Diarrhea 3	4/6	6.49 <sup>0.9</sup> (2.48)	2.33 <sup>0.5</sup> (0.89)	3.15 <sup>0.8</sup> (1.39)	7 <sup>0.9</sup> (3.28)	2.33 <sup>0.4</sup> (0.82)	2.44 <sup>0.4</sup> (0.82)	6.08 <sup>0.9</sup> (2.7)	
Asthenia 1	41/368	0.66 (0.17)	0.52 (0.11)	0.63 (0.14)	1.01 <sup>0.1</sup> (0.7)	0.54 (0.12)	0.66 (0.18)	1.05 <sup>0.1</sup> (0.54)	
Asthenia 2	32/150	0.83 (0.18)	0.75 (0.19)	0.84 (0.34)	1.31 <sup>0.2</sup> (0.88)	0.78 (0.21)	1 (0.35)	1.74 <sup>0.3</sup> (1.3)	
Asthenia 3	4/21	2.26 <sup>0.4</sup> (0.8)	1.03 (0.19)	1.4 (0.32)	3.73 <sup>0.8</sup> (1.48)	1.02 (0.15)	1.48 <sup>0.1</sup> (0.35)	4.4 <sup>0.7</sup> (1.96)	
Neutropenia 1	12/111	0.65 (0.15)	0.41 (0.1)	0.43 (0.13)	0.62 (0.16)	0.4 (0.09)	0.37 (0.11)	0.56 (0.15)	
Neutropenia 2	9/67	0.65 (0.05)	0.47 (0.08)	0.51 (0.13)	0.65 (0.18)	0.44 (0.07)	0.43 (0.07)	0.57 (0.1)	
Neutropenia 3	4/9	3.73 <sup>0.4</sup> (2.12)	1.84 <sup>0.4</sup> (0.84)	2.22 <sup>0.4</sup> (1.2)	3.47 <sup>0.7</sup> (1.92)	1.74 <sup>0.3</sup> (0.67)	2.18 <sup>0.3</sup> (1.03)	4.16 <sup>0.8</sup> (1.64)	
Neutropenia 4	1/5	1.81 (0.48)	1.07 (0.23)	1.37 <sup>0.1</sup> (0.64)	2.64 <sup>0.5</sup> (1.64)	1.04 (0.18)	1.31 (0.42)	2.78 <sup>0.4</sup> (1.22)	
Thrombopenia 1	15/104	1.49 <sup>0.1</sup> (0.61)	0.58 (0.14)	0.63 (0.15)	1.24 <sup>0.1</sup> (0.55)	0.57 (0.15)	0.58 (0.15)	1.4 <sup>0.1</sup> (0.78)	
Thrombopenia 2	2/6	3.22 <sup>0.3</sup> (1.35)	0.88 (0.25)	0.74 <sup>0.1</sup> (0.55)	0.69 (0.29)	0.82 (0.2)	0.8 (0.41)	1.41 <sup>0.2</sup> (1.39)	
Thrombopenia 3	0/1	3.21 (0.52)	1.45 (0.12)	2.27 <sup>0.3</sup> (0.88)	4.79 <sup>0.8</sup> (1.97)	1.34 (0.11)	1.52 (0.14)	4.07 <sup>0.4</sup> (2.38)	
Thrombopenia 4	0/1	1.13 (0.11)	0.92 (0.05)	1.08 (0.34)	1.58 <sup>0.3</sup> (1.14)	0.9 (0.04)	1.14 (0.3)	0.95 (0.17)	
Anemia 1	26/257	0.42 (0.09)	0.2 (0.04)	0.2 (0.06)	0.31 (0.11)	0.19 (0.03)	0.18 (0.05)	0.28 (0.09)	
Anemia 2	12/36	3.25 <sup>0.8</sup> (1.09)	1.24 (0.25)	1.03 (0.22)	1.25 <sup>0.1</sup> (0.46)	1.21 (0.21)	1.02 (0.12)	1.35 <sup>0.1</sup> (0.33)	
Number of selected variables	-	5	3	6	11	2	4	9	
AUC	-	0.69 (0.04)	0.69 (0.03)	0.69 (0.04)	0.68 (0.04)	0.68 (0.03)	0.68 (0.04)	0.68 (0.03)	

Note: <sup>s</sup>,  $0 \leq s \leq 1$  is the proportion of times a variable is selected by the model. The \* column indicates the proportion of dose reductions linked to the variable.



**Table 8.** Estimated coefficients, number of selected variables and AUCs for usual SSVS and SSVS using the power prior for the second approach with  $P_0 = P$ ,  $a_0 \in \{0.1, 0.5, 1\}$ , BMA and several values of  $n_0$ .

Variables / Models	*	Power prior ( $P_0 = P$ )				$n_0 = n_1$		$n_0 = 1.77 \times n_1$	
		$a_0 = 0.1$	$a_0 = 0.5$	$a_0 = 1$	$a_0 = 0.1$	$a_0 = 0.5$	$a_0 = 0.1$	$a_0 = 0.5$	
Intercept	-	-10.62 (1.36)	-5.27 (0.48)	-5.61 (0.51)	-10.3 (1.76)	-5.05 (0.51)	-5.79 (0.93)	-11.65 (1.29)	
Age $\geq 80$ years	5/32	1.63 <sup>0.1</sup> (0.68)	1.4 <sup>0.1</sup> (0.7)	2.99 <sup>0.5</sup> (1.65)	6.14 <sup>1</sup> (2.63)	1.69 <sup>0.2</sup> (0.94)	3.93 <sup>0.8</sup> (2.13)	8.86 <sup>1</sup> (4.47)	
Weight loss $> 10\%$	0/5	1.69 (0.64)	0.87 (0.18)	1.06 <sup>0.1</sup> (0.57)	2.17 <sup>0.3</sup> (2.19)	0.83 (0.12)	1 <sup>0.1</sup> (0.4)	2.19 <sup>0.2</sup> (2.02)	
ECOG performance status 1	47/381	0.6 (0.23)	0.56 (0.19)	0.62 (0.19)	0.81 (0.29)	0.55 (0.19)	0.61 (0.24)	0.94 <sup>0.1</sup> (0.59)	
ECOG performance status 2	25/100	1.87 <sup>0.1</sup> (0.48)	1.04 (0.27)	0.89 (0.52)	1.84 <sup>0.4</sup> (1.71)	1.1 (0.37)	1.07 <sup>0.1</sup> (0.71)	2.16 <sup>0.3</sup> (2.3)	
ECOG performance status 3	2/11	1.61 (0.38)	0.91 (0.17)	1.42 <sup>0.1</sup> (0.59)	4.12 <sup>0.7</sup> (2.01)	0.92 (0.18)	1.53 <sup>0.1</sup> (0.57)	4.7 <sup>0.8</sup> (2.17)	
ECOG performance status 4	3/21	1.24 (0.49)	1.05 (0.29)	2.34 <sup>0.4</sup> (1.31)	6.05 <sup>1</sup> (2.11)	1.02 (0.31)	2.52 <sup>0.4</sup> (1.23)	7.45 <sup>1</sup> (1.58)	
Bilirubin $> 35 \mu\text{mol/L}$	1/1	5.94 <sup>1</sup> (0.38)	4.37 <sup>0.6</sup> (1.75)	5.15 <sup>0.8</sup> (3.04)	8.35 <sup>0.8</sup> (4.18)	6.97 <sup>0.9</sup> (1.14)	9.13 <sup>0.9</sup> (2.84)	13.09 <sup>0.8</sup> (6.47)	
Bilirubin $> 50 \mu\text{mol/L}$	2/5	4.39 <sup>0.7</sup> (1.64)	2.36 <sup>0.4</sup> (1.07)	3.46 <sup>0.7</sup> (1.73)	7.41 <sup>0.9</sup> (3.1)	2.22 <sup>0.4</sup> (1.06)	3.75 <sup>0.7</sup> (2.18)	8.44 <sup>1</sup> (3.68)	
Treatment line 3	8/46	1.28 (0.35)	0.62 (0.19)	0.52 (0.16)	0.63 (0.14)	0.59 (0.21)	0.51 (0.2)	0.77 (0.21)	
Treatment line $> 3$	6/43	1.76 (0.6)	0.76 (0.22)	0.79 (0.42)	1.58 <sup>0.2</sup> (1.45)	0.77 (0.18)	0.96 (0.5)	2.36 <sup>0.3</sup> (2.35)	
Vomiting 1	11/54	1.33 <sup>0.1</sup> (0.52)	0.7 (0.14)	0.7 (0.19)	0.99 (0.26)	0.69 (0.16)	0.76 (0.21)	1.21 (0.34)	
Vomiting 2	5/13	2.21 <sup>0.4</sup> (0.95)	0.92 (0.3)	0.91 (0.47)	1.53 <sup>0.3</sup> (0.98)	0.84 (0.26)	0.88 (0.48)	1.81 <sup>0.3</sup> (1.53)	
Vomiting 3	1/2	3.69 <sup>0.4</sup> (1.86)	2.12 <sup>0.2</sup> (1.78)	2.25 <sup>0.1</sup> (2.39)	4.9 <sup>0.7</sup> (3.09)	1.66 <sup>0.1</sup> (1.06)	2.32 <sup>0.2</sup> (2.55)	5.05 <sup>0.4</sup> (5.04)	
Nausea 1	28/226	0.62 (0.13)	0.31 (0.06)	0.24 (0.05)	0.33 (0.09)	0.29 (0.06)	0.26 (0.09)	0.41 (0.12)	
Nausea 2	13/60	0.63 (0.12)	0.47 (0.08)	0.5 (0.11)	0.74 (0.46)	0.46 (0.08)	0.47 (0.14)	0.61 (0.19)	
Nausea 3	2/6	2.21 <sup>0.2</sup> (2.27)	0.93 (0.4)	1 (0.34)	2.55 <sup>0.3</sup> (2.31)	0.87 (0.32)	1 <sup>0.1</sup> (0.65)	1.87 <sup>0.1</sup> (2.26)	
Diarrhea 1	28/218	0.48 (0.11)	0.3 (0.09)	0.26 (0.1)	0.35 (0.11)	0.27 (0.08)	0.23 (0.09)	0.29 (0.1)	
Diarrhea 2	18/49	6.54 <sup>1</sup> (1.88)	2.22 <sup>0.5</sup> (0.63)	2.09 <sup>0.4</sup> (0.62)	4.07 <sup>0.9</sup> (1.32)	2.05 <sup>0.4</sup> (0.57)	2.18 <sup>0.4</sup> (0.69)	4.95 <sup>1</sup> (0.93)	
Diarrhea 3	4/6	6.49 <sup>0.9</sup> (2.48)	3.2 <sup>0.7</sup> (1.03)	4.07 <sup>1</sup> (0.93)	7.92 <sup>1</sup> (1.64)	3.12 <sup>0.7</sup> (1.1)	4.17 <sup>0.9</sup> (1.46)	8.22 <sup>1</sup> (3.01)	
Asthenia 1	41/368	0.66 (0.17)	0.52 (0.11)	0.61 (0.12)	0.93 (0.25)	0.51 (0.12)	0.71 (0.25)	1.35 <sup>0.2</sup> (0.88)	
Asthenia 2	32/150	0.83 (0.18)	0.75 (0.21)	0.89 (0.31)	1.68 <sup>0.2</sup> (1.03)	0.71 (0.21)	0.83 (0.35)	1.83 <sup>0.3</sup> (1.48)	
Asthenia 3	4/21	2.26 <sup>0.4</sup> (0.8)	1.09 (0.23)	1.52 <sup>0.1</sup> (0.48)	4.15 <sup>0.9</sup> (1.71)	1.04 (0.23)	1.71 <sup>0.2</sup> (0.57)	5.15 <sup>0.9</sup> (1.75)	
Neutropenia 1	12/111	0.65 (0.15)	0.39 (0.09)	0.37 (0.12)	0.5 (0.16)	0.37 (0.08)	0.37 (0.08)	0.55 (0.16)	
Neutropenia 2	9/67	0.65 (0.05)	0.46 (0.08)	0.43 (0.1)	0.56 (0.17)	0.44 (0.07)	0.4 (0.08)	0.57 (0.31)	
Neutropenia 3	4/9	3.73 <sup>0.4</sup> (2.12)	1.75 <sup>0.3</sup> (0.91)	1.78 <sup>0.2</sup> (1.21)	2.91 <sup>0.3</sup> (2.97)	1.64 <sup>0.3</sup> (0.74)	1.66 <sup>0.3</sup> (0.74)	3.52 <sup>0.7</sup> (1.8)	
Neutropenia 4	1/5	1.81 (0.48)	1.06 (0.23)	1.4 <sup>0.1</sup> (0.92)	2.73 <sup>0.4</sup> (2.17)	1.04 (0.18)	1.31 (0.37)	3.23 <sup>0.7</sup> (1.62)	
Thrombopenia 1	15/104	1.49 <sup>0.1</sup> (0.61)	0.54 (0.14)	0.5 (0.13)	0.95 (0.42)	0.5 (0.11)	0.47 (0.14)	0.97 (0.31)	
Thrombopenia 2	2/6	3.22 <sup>0.3</sup> (1.35)	1.06 <sup>0.1</sup> (0.73)	0.65 (0.49)	1 <sup>0.1</sup> (1.41)	0.9 (0.46)	0.53 (0.21)	1 <sup>0.1</sup> (1.28)	
Thrombopenia 3	0/1	3.21 (0.52)	1.42 (0.17)	1.5 <sup>0.1</sup> (0.65)	2.87 <sup>0.3</sup> (2.71)	1.27 (0.11)	1.3 (0.23)	3.58 <sup>0.4</sup> (3.16)	
Thrombopenia 4	0/1	1.13 (0.11)	0.89 (0.05)	0.87 (0.08)	0.92 (0.05)	0.91 (0.04)	0.93 (0.11)	1.02 (0.2)	
Anemia 1	26/257	0.42 (0.09)	0.2 (0.04)	0.17 (0.04)	0.26 (0.09)	0.18 (0.03)	0.17 (0.03)	0.26 (0.08)	
Anemia 2	12/36	3.25 <sup>0.8</sup> (1.09)	1.21 (0.26)	0.9 (0.18)	1.14 (0.43)	1.17 (0.23)	0.86 (0.16)	1.28 <sup>0.1</sup> (0.8)	
Number of selected variables	-	5	3	4	9	2	4	10	
AUC	-	0.69 (0.04)	0.68 (0.03)	0.68 (0.04)	0.69 (0.05)	0.6 (0.03)	0.68 (0.05)	0.68 (0.05)	

Note: <sup>s</sup>,  $0 \leq s \leq 1$  is the proportion of times a variable is selected by the model. The \* column indicates the proportion of dose reductions linked to the variable.



**Table 9.** Estimated coefficients, number of selected variables and AUCs for usual SSVS and SSVS using the power prior for the second approach with  $P_0 = 2P$ ,  $a_0 \in \{0.1, 0.5, 1\}$ , BMA and several values of  $n_0$ .

Variables / Models	*	Power prior ( $P_0 = 2P$ )						$n_0 = 1.77 \times n_1$ $a_0 = 1$
		SSVS		$a_0 = 0.1$		$a_0 = 0.5$		
		$n_0 = n_1$ $a_0 = 1$	$a_0 = 0.1$	$a_0 = 0.5$	$a_0 = 0.1$	$a_0 = 0.5$	$a_0 = 0.1$	
Intercept	-	-10.62 (1.36)	-4.66 (0.37)	-5.36 (0.7)	-11.15 (2.29)	-4.5 (0.39)	-5.28 (0.72)	-11.43 (1.84)
Age $\geq 80$ years	5/32	1.63 <sup>0.1</sup> (0.68)	1.33 <sup>0.1</sup> (0.52)	3.57 <sup>0.6</sup> (1.95)	8.82 <sup>1</sup> (2.64)	1.46 <sup>0.2</sup> (0.81)	3.77 <sup>0.7</sup> (2.01)	9.85 <sup>1</sup> (2.74)
Weight loss $> 10\%$	0/5	1.69 (0.64)	1.1 (0.41)	1.69 <sup>0.3</sup> (0.88)	3.87 <sup>0.6</sup> (2.83)	1.04 (0.28)	1.68 <sup>0.3</sup> (0.66)	3.78 <sup>0.6</sup> (2.11)
ECOG performance status 1	47/381	0.6 (0.23)	0.52 (0.19)	0.54 (0.2)	0.68 (0.2)	0.5 (0.17)	0.54 (0.21)	0.75 (0.45)
ECOG performance status 2	25/100	1.87 <sup>0.1</sup> (0.48)	1.23 <sup>0.1</sup> (0.43)	1.23 <sup>0.1</sup> (0.89)	2.99 <sup>0.4</sup> (3.04)	1.22 (0.4)	1.32 <sup>0.2</sup> (0.89)	3.29 <sup>0.4</sup> (3.11)
ECOG performance status 3	2/11	1.61 (0.38)	1.01 (0.23)	2.58 <sup>0.5</sup> (1.12)	7.71 <sup>1</sup> (1.5)	0.99 (0.22)	2.84 <sup>0.5</sup> (1.65)	8.7 <sup>1</sup> (2.02)
ECOG performance status 4	3/21	1.24 (0.49)	0.96 (0.2)	2.24 <sup>0.6</sup> (0.87)	6.53 <sup>0.9</sup> (2.25)	0.95 (0.26)	2.21 <sup>0.3</sup> (0.96)	7.13 <sup>1</sup> (1.49)
Bilirubin $> 35 \mu\text{mol/L}$	1/1	5.94 <sup>1</sup> (0.38)	6.29 <sup>1</sup> (0.79)	6.96 <sup>1</sup> (2.43)	10.22 <sup>1</sup> (1.07)	6.29 <sup>1</sup> (1.05)	8.66 <sup>1</sup> (2.11)	13.46 <sup>1</sup> (1.79)
Bilirubin $> 50 \mu\text{mol/L}$	2/5	4.39 <sup>0.7</sup> (1.64)	2.14 <sup>0.3</sup> (0.99)	3.61 <sup>0.6</sup> (2.33)	8.64 <sup>1</sup> (2.89)	2.02 <sup>0.4</sup> (0.97)	3.81 <sup>0.7</sup> (2.08)	9.55 <sup>1</sup> (3.13)
Treatment line 3	8/46	1.28 (0.35)	0.56 (0.2)	0.49 (0.21)	0.6 (0.18)	0.53 (0.22)	0.44 (0.21)	0.55 (0.21)
Treatment line $> 3$	6/43	1.76 (0.6)	0.74 (0.2)	1.01 (0.53)	3.32 <sup>0.4</sup> (3.09)	0.69 (0.16)	0.92 (0.46)	3.01 <sup>0.4</sup> (2.7)
Vomiting 1	11/54	1.33 <sup>0.1</sup> (0.52)	0.63 (0.16)	0.67 (0.26)	0.99 (0.32)	0.61 (0.15)	0.7 (0.29)	1.09 <sup>0.1</sup> (0.47)
Vomiting 2	5/13	2.21 <sup>0.4</sup> (0.95)	1.09 (0.49)	1.77 <sup>0.4</sup> (1.32)	3.24 <sup>0.4</sup> (2.9)	0.9 (0.31)	1.04 <sup>0.1</sup> (1.15)	1.76 <sup>0.2</sup> (2.12)
Vomiting 3	1/2	3.69 <sup>0.4</sup> (1.86)	1.92 <sup>0.1</sup> (1.43)	2.59 <sup>0.1</sup> (2.11)	5.7 <sup>0.8</sup> (2.04)	2.02 <sup>0.2</sup> (1.8)	2.23 <sup>0.2</sup> (1.42)	5.23 <sup>0.5</sup> (2.46)
Nausea 1	28/226	0.62 (0.13)	0.28 (0.06)	0.27 (0.11)	0.39 (0.11)	0.27 (0.07)	0.25 (0.12)	0.39 (0.2)
Nausea 2	13/60	0.63 (0.12)	0.43 (0.06)	0.45 (0.18)	0.79 (0.6)	0.42 (0.07)	0.36 (0.11)	0.46 (0.16)
Nausea 3	2/6	2.21 <sup>0.2</sup> (2.27)	0.9 (0.43)	1.13 (0.57)	2.85 <sup>0.4</sup> (2)	0.93 <sup>0.1</sup> (0.51)	1.13 (0.76)	2.31 <sup>0.4</sup> (1.47)
Diarrhea 1	28/218	0.48 (0.11)	0.3 (0.09)	0.31 (0.11)	0.43 (0.16)	0.28 (0.09)	0.27 (0.13)	0.34 (0.13)
Diarrhea 2	18/49	6.54 <sup>1</sup> (1.88)	1.82 <sup>0.3</sup> (0.55)	2.12 <sup>0.4</sup> (1.14)	5.05 <sup>1</sup> (1.74)	1.76 <sup>0.2</sup> (0.52)	2.41 <sup>0.5</sup> (1.02)	6.26 <sup>1</sup> (0.98)
Diarrhea 3	4/6	6.49 <sup>0.9</sup> (2.48)	4.18 <sup>0.9</sup> (1.44)	6.87 <sup>1</sup> (2.66)	13.31 <sup>1</sup> (4.02)	3.94 <sup>0.9</sup> (1.64)	6.88 <sup>1</sup> (3.08)	13.93 <sup>1</sup> (4.92)
Asthenia 1	41/368	0.66 (0.17)	0.42 (0.07)	0.52 (0.13)	0.9 (0.3)	0.41 (0.08)	0.53 (0.15)	0.93 (0.4)
Asthenia 2	32/150	0.83 (0.18)	0.67 (0.22)	1.02 (0.56)	2.74 <sup>0.4</sup> (2.23)	0.65 (0.2)	0.94 (0.53)	2.53 <sup>0.4</sup> (2.19)
Asthenia 3	4/21	2.26 <sup>0.4</sup> (0.8)	1.02 (0.23)	2.36 <sup>0.3</sup> (1.17)	7.29 <sup>1</sup> (1.85)	0.98 (0.28)	2.48 <sup>0.4</sup> (1.28)	7.41 <sup>1</sup> (1.21)
Neutropenia 1	12/111	0.65 (0.15)	0.35 (0.08)	0.34 (0.08)	0.58 (0.19)	0.33 (0.07)	0.32 (0.08)	0.54 (0.25)
Neutropenia 2	9/67	0.65 (0.05)	0.43 (0.07)	0.41 (0.1)	0.56 (0.22)	0.4 (0.06)	0.33 (0.06)	0.46 (0.1)
Neutropenia 3	4/9	3.73 <sup>0.4</sup> (2.12)	1.73 <sup>0.3</sup> (1)	1.96 <sup>0.3</sup> (1.4)	3.76 <sup>0.3</sup> (3.19)	1.62 <sup>0.3</sup> (0.77)	1.96 <sup>0.3</sup> (1.18)	4.04 <sup>0.6</sup> (2.92)
Neutropenia 4	1/5	1.81 (0.48)	1.03 (0.23)	1.69 <sup>0.1</sup> (1.06)	4.8 <sup>0.7</sup> (2.52)	1.06 (0.21)	1.81 <sup>0.2</sup> (0.53)	6.03 <sup>1</sup> (1.86)
Thrombopenia 1	15/104	1.49 <sup>0.1</sup> (0.61)	0.43 (0.09)	0.39 (0.1)	0.78 (0.24)	0.4 (0.09)	0.37 (0.1)	0.83 (0.31)
Thrombopenia 2	2/6	3.22 <sup>0.3</sup> (1.35)	1.22 <sup>0.1</sup> (1.01)	0.55 (0.15)	0.62 (0.2)	1.17 <sup>0.1</sup> (0.95)	0.54 (0.14)	0.66 (0.28)
Thrombopenia 3	0/1	3.21 (0.52)	1.46 (0.19)	2.46 <sup>0.3</sup> (1.26)	6.22 <sup>0.7</sup> (3.69)	1.32 (0.18)	2.21 <sup>0.5</sup> (0.97)	4.73 <sup>0.7</sup> (2.6)
Thrombopenia 4	0/1	1.13 (0.11)	0.92 (0.07)	1.01 (0.26)	1.4 (0.54)	0.93 (0.06)	0.97 (0.15)	1.23 (0.3)
Anemia 1	26/257	0.42 (0.09)	0.17 (0.03)	0.17 (0.04)	0.27 (0.09)	0.16 (0.03)	0.17 (0.03)	0.3 (0.08)
Anemia 2	12/36	3.25 <sup>0.8</sup> (1.09)	1.1 (0.26)	0.8 (0.25)	0.99 (0.43)	1.08 (0.21)	0.77 (0.15)	0.88 (0.27)
Number of selected variables	-	5	2	6	13	2	7	13
AUC	-	0.69 (0.04)	0.69 (0.03)	0.68 (0.04)	0.69 (0.05)	0.69 (0.03)	0.69 (0.05)	0.69 (0.05)

Note: <sup>s</sup>,  $0 \leq s \leq 1$  is the proportion of times a variable is selected by the model. The \* column indicates the proportion of dose reductions linked to the variable.

**Table 10.** Estimated coefficients, number of selected variables and AUCs for usual SSVS and SSVS using the power prior for the second approach with  $P_0 = P/2$ ,  $\alpha_0 \in \{0.1, 0.5, 1\}$ , BMA and several values of  $n_0$ .

Variables / Models	*	Power prior ( $P_0 = P/2$ )				$n_0 = n_1$		$n_0 = 1.77 \times n_1$	
		$\alpha_0 = 0.1$	$\alpha_0 = 0.5$	$\alpha_0 = 1$	$\alpha_0 = 0.1$	$\alpha_0 = 0.5$	$\alpha_0 = 0.1$	$\alpha_0 = 0.5$	
Intercept	-	-10.62 (1.36)	-5.49 (0.43)	-5.98 (0.65)	-10.92 (1.5)	-5.3 (0.43)	-5.59 (0.55)	-11.35 (1.65)	
Age $\geq 80$ years	5/32	1.63 <sup>0.1</sup> (0.68)	1.15 <sup>0.1</sup> (0.53)	2.49 <sup>0.4</sup> (1.66)	4.99 <sup>0.9</sup> (3.17)	1.18 (0.43)	2.38 <sup>0.3</sup> (1.08)	6.31 <sup>1</sup> (1.79)	
Weight loss > 10%	0/5	1.69 (0.64)	0.8 (0.15)	0.79 (0.15)	1.31 (0.44)	0.82 (0.14)	0.83 (0.16)	1.65 <sup>0.2</sup> (0.76)	
ECOG performance status 1	47/381	0.6 (0.23)	0.57 (0.18)	0.63 (0.18)	0.75 (0.29)	0.55 (0.17)	0.56 (0.2)	0.73 (0.37)	
ECOG performance status 2	25/100	1.87 <sup>0.1</sup> (0.48)	0.94 (0.19)	0.85 (0.41)	1.46 <sup>0.3</sup> (1.32)	0.89 (0.19)	0.68 (0.33)	1.22 <sup>0.3</sup> (0.92)	
ECOG performance status 3	2/11	1.61 (0.38)	0.91 (0.16)	1.08 (0.33)	3.38 <sup>0.6</sup> (1.43)	0.88 (0.12)	1.09 <sup>0.3</sup> (0.43)	3.19 <sup>0.6</sup> (1.8)	
ECOG performance status 4	3/21	1.24 (0.49)	1.04 (0.29)	2.12 <sup>0.4</sup> (1.38)	6.27 <sup>0.8</sup> (3.7)	1.06 (0.24)	1.92 <sup>0.2</sup> (0.53)	6.61 <sup>0.2</sup> (2.65)	
Bilirubin > 35 $\mu\text{mol/L}$	1/1	5.94 <sup>1</sup> (0.38)	3.86 <sup>0.4</sup> (2.3)	4.74 <sup>0.7</sup> (3.83)	8.57 <sup>0.9</sup> (4.81)	3.38 <sup>0.5</sup> (0.62)	2.87 <sup>0.5</sup> (1.88)	6.59 <sup>0.7</sup> (3.67)	
Bilirubin > 50 $\mu\text{mol/L}$	2/5	4.39 <sup>0.7</sup> (1.64)	2.33 <sup>0.4</sup> (0.92)	3.3 <sup>0.6</sup> (2.08)	7.56 <sup>0.9</sup> (3.74)	2.27 <sup>0.5</sup> (0.87)	3.17 <sup>0.6</sup> (1.54)	8.22 <sup>1</sup> (3.23)	
Treatment line 3	8/46	1.28 (0.35)	0.66 (0.2)	0.57 (0.2)	0.68 (0.22)	0.63 (0.21)	0.52 (0.21)	0.75 (0.22)	
Treatment line > 3	6/43	1.76 (0.6)	0.75 (0.19)	0.77 (0.25)	1.26 <sup>0.1</sup> (0.73)	0.69 (0.16)	0.6 (0.13)	0.98 (0.39)	
Vomiting 1	11/54	1.33 <sup>0.1</sup> (0.52)	0.73 (0.15)	0.74 (0.19)	1.1 (0.39)	0.74 (0.15)	0.79 (0.15)	1.55 <sup>0.2</sup> (0.48)	
Vomiting 2	5/13	2.21 <sup>0.4</sup> (0.95)	0.92 (0.28)	0.83 (0.38)	1.64 <sup>0.3</sup> (1.12)	0.83 (0.25)	0.69 (0.25)	1.25 <sup>0.1</sup> (0.78)	
Vomiting 3	1/2	3.69 <sup>0.4</sup> (1.86)	2.07 <sup>0.2</sup> (1.54)	1.62 <sup>0.1</sup> (1.33)	3.17 <sup>0.3</sup> (2.33)	2.01 <sup>0.2</sup> (1.52)	2.21 <sup>0.2</sup> (2.26)	4.85 <sup>0.4</sup> (3.73)	
Nausea 1	28/226	0.62 (0.13)	0.32 (0.05)	0.28 (0.05)	0.41 (0.08)	0.29 (0.05)	0.25 (0.06)	0.36 (0.08)	
Nausea 2	13/60	0.63 (0.12)	0.49 (0.08)	0.48 (0.07)	0.65 (0.17)	0.47 (0.07)	0.48 (0.09)	0.62 (0.13)	
Nausea 3	2/6	2.21 <sup>0.2</sup> (2.27)	0.95 (0.43)	0.9 (0.27)	1.34 <sup>0.1</sup> (0.67)	0.91 (0.4)	0.86 (0.25)	2.11 <sup>0.3</sup> (2.25)	
Diarrhea 1	28/218	0.48 (0.11)	0.29 (0.09)	0.26 (0.1)	0.36 (0.11)	0.27 (0.08)	0.23 (0.07)	0.33 (0.09)	
Diarrhea 2	18/49	6.54 <sup>1</sup> (1.88)	2.4 <sup>0.7</sup> (0.6)	2.42 <sup>0.6</sup> (0.63)	4.95 <sup>1</sup> (0.88)	2.27 <sup>0.6</sup> (0.6)	2.16 <sup>0.5</sup> (0.63)	5.2 <sup>1</sup> (1.31)	
Diarrhea 3	4/6	6.49 <sup>0.9</sup> (2.48)	2.31 <sup>0.5</sup> (0.99)	3.01 <sup>0.7</sup> (1.32)	6.44 <sup>0.9</sup> (2.36)	2.14 <sup>0.4</sup> (0.67)	2.61 <sup>0.7</sup> (1.18)	6.15 <sup>0.9</sup> (2.58)	
Asthenia 1	41/368	0.66 (0.17)	0.55 (0.1)	0.64 (0.1)	0.77 (0.17)	0.56 (0.12)	0.67 (0.18)	0.88 (0.28)	
Asthenia 2	32/150	0.83 (0.18)	0.78 (0.17)	0.86 (0.25)	1.13 <sup>0.1</sup> (0.5)	0.79 (0.16)	0.93 (0.24)	1.24 (0.48)	
Asthenia 3	4/21	2.26 <sup>0.4</sup> (0.8)	1.09 (0.23)	1.46 (0.32)	3.52 <sup>0.8</sup> (1.27)	1.05 (0.19)	1.42 (0.27)	3.82 <sup>0.8</sup> (1.31)	
Neutropenia 1	12/111	0.65 (0.15)	0.42 (0.11)	0.41 (0.14)	0.59 (0.2)	0.39 (0.09)	0.35 (0.09)	0.51 (0.19)	
Neutropenia 2	9/67	0.65 (0.05)	0.47 (0.08)	0.45 (0.12)	0.55 (0.2)	0.44 (0.07)	0.42 (0.07)	0.6 (0.08)	
Neutropenia 3	4/9	3.73 <sup>0.4</sup> (2.12)	1.81 <sup>0.4</sup> (0.81)	1.84 <sup>0.3</sup> (1.02)	2.71 <sup>0.5</sup> (1.57)	1.7 <sup>0.3</sup> (0.68)	1.64 <sup>0.3</sup> (0.84)	2.83 <sup>0.5</sup> (1.76)	
Neutropenia 4	1/5	1.81 (0.48)	1.07 (0.19)	1.13 (0.35)	2.04 <sup>0.3</sup> (1.25)	1.03 (0.16)	1.14 (0.17)	2.65 <sup>0.3</sup> (1.6)	
Thrombopenia 1	15/104	1.49 <sup>0.1</sup> (0.61)	0.61 (0.17)	0.62 (0.16)	1.25 (0.49)	0.57 (0.14)	0.55 (0.12)	1.33 (0.59)	
Thrombopenia 2	2/6	3.22 <sup>0.3</sup> (1.35)	0.88 (0.26)	0.71 <sup>0.1</sup> (0.54)	0.69 (0.21)	0.75 (0.17)	0.49 (0.15)	0.55 (0.12)	
Thrombopenia 3	0/1	3.21 (0.52)	1.37 (0.13)	1.35 (0.39)	2.41 <sup>0.5</sup> (1.22)	1.27 (0.19)	1.07 (0.21)	2 (0.77)	
Thrombopenia 4	0/1	1.13 (0.11)	0.9 (0.04)	0.89 (0.07)	1.03 (0.2)	0.9 (0.04)	0.99 (0.2)	0.94 (0.1)	
Anemia 1	26/257	0.42 (0.09)	0.2 (0.04)	0.18 (0.05)	0.3 (0.1)	0.19 (0.03)	0.16 (0.03)	0.29 (0.09)	
Anemia 2	12/36	3.25 <sup>0.8</sup> (1.09)	1.29 (0.26)	1.02 (0.2)	1.16 (0.24)	1.24 (0.24)	0.96 (0.13)	1.29 (0.34)	
Number of selected variables	-	5	2	4	10	3	4	9	
AUC	-	0.69 (0.04)	0.69 (0.03)	0.68 (0.04)	0.69 (0.04)	0.69 (0.03)	0.68 (0.03)	0.68 (0.03)	

Note: <sup>s</sup>,  $0 \leq s \leq 1$  is the proportion of times a variable is selected by the model. The \* column indicates the proportion of dose reductions linked to the variable.

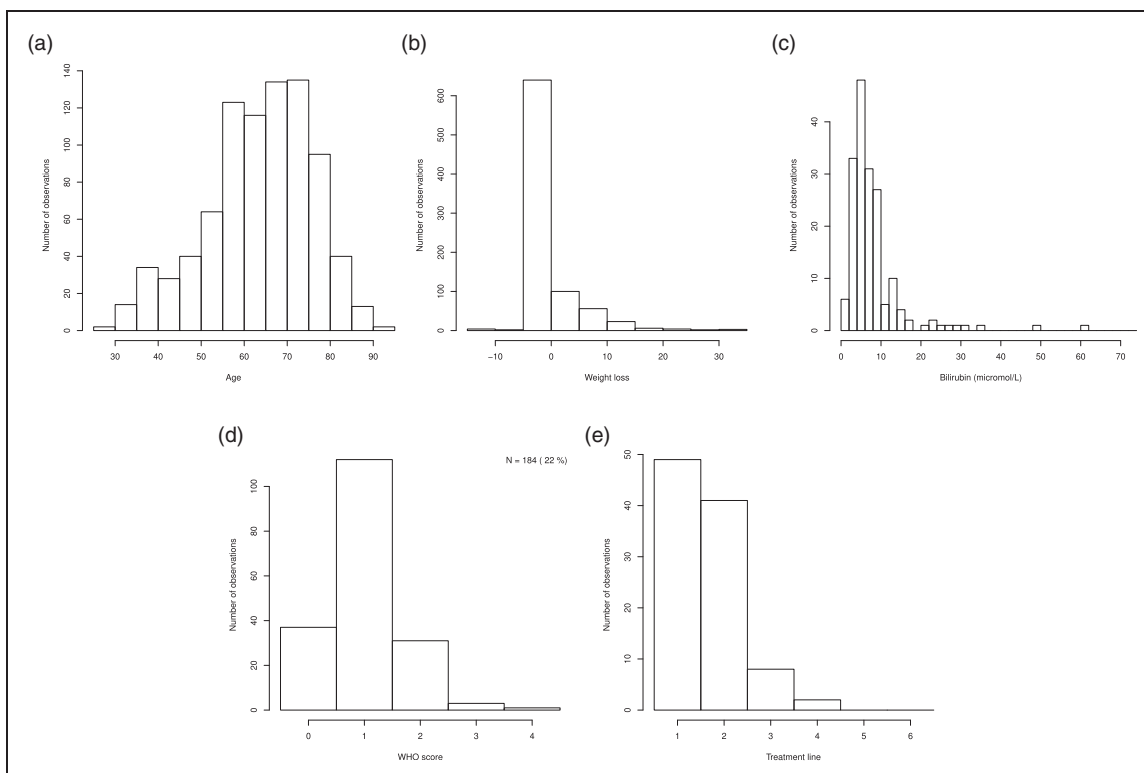
## Appendix 2

Among the 678 included patients,  $n_0 = 375$  have incomplete observations (sample 2; see Table 11, Figures 9 to 11 for description). Among patients with incomplete observations over their cycles, 171 have dose reductions.

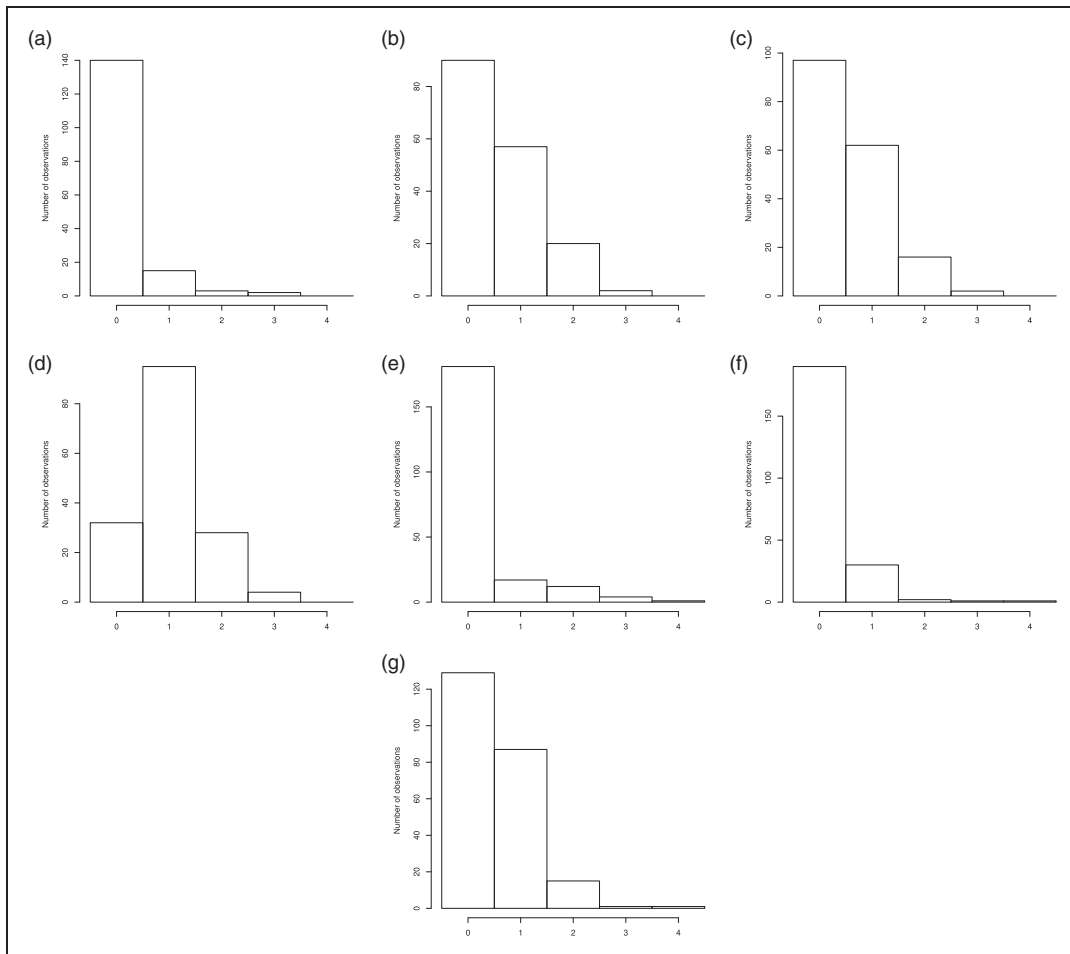
**Table 11.** Description of covariates for the  $n_0$  patients with incomplete observations (sample 2).

Variables	NA	0	1	2	3	4
Age $\geq 80$ years	–	780 (93)	60 (7)	–	–	–
Weight loss $> 10\%$	–	802 (95)	38 (5)	–	–	–
ECOG performance status	656 (78)	37 (20)	112 (61)	31 (17)	3 (2)	1 (1)
Bilirubin $> 35, > 50 \mu\text{mol/L}$	661 (79)	173 (97)	1 (1)	5 (3)	–	–
Treatment line 3, $> 3$	740 (88)	90 (90)	8 (8)	2 (2)	–	–
<b>Toxicity grades</b>						
Vomiting	680 (81)	140 (88)	15 (9)	3 (2)	2 (1)	–
Nausea	671 (80)	90 (53)	57 (34)	20 (12)	2 (1)	–
Diarrhea	663 (79)	97 (55)	62 (35)	16 (9)	2 (1)	–
Asthenia	681 (81)	32 (20)	95 (60)	28 (18)	4 (3)	–
Neutropenia	625 (74)	181 (84)	17 (8)	12 (6)	4 (2)	1 (0)
Thrombopenia	616 (73)	190 (85)	30 (13)	2 (1)	1 (0)	1 (0)
Anemia	607 (72)	129 (55)	87 (37)	15 (6)	1 (0)	1 (0)

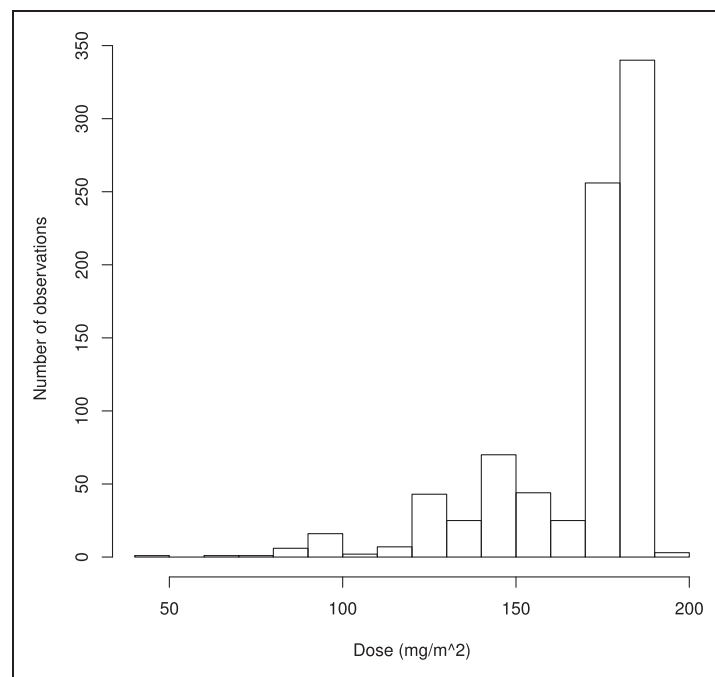
Note: Percentages are shown in parentheses.



**Figure 9.** Distribution of age, weight loss, bilirubin, ECOG performance status and treatment line for the  $n_0$  patients with incomplete observations (sample 2). (a) Age; (b) weight loss; (c) Bilirubin; (d) ECOG performance status; and (e) treatment line.



**Figure 10.** Distribution of toxicities for the  $n_0$  patients with incomplete observations (sample 2). (a) Vomiting; (b) nausea; (c) diarrhea; (d) asthenia; (e) neutropenia; (f) thrombopenia; and (g) anemia.



**Figure 11.** Distribution of doses for the  $n_0$  patients with incomplete observations (sample 2).



# Synthèse

## Contexte

Aujourd'hui, la médecine de précision autorise la personnalisation des soins de santé au travers de décisions, de pratiques et/ou de produits médicaux adaptés à chaque patient. La médecine de précision inclut la sélection et l'optimisation du traitement et des doses afin de garantir les meilleurs bénéfices pour un sous-groupe de patients donné. Les stratégies de traitement passent donc de l'approche «taille unique» à des dosages médicamenteux plus individualisés. Identifier le bon médicament pour un sous-groupe donné n'est donc qu'une partie du traitement : donner la bonne dose au bon moment est tout aussi important.

En oncologie, la prise en charge de patients atteints d'un cancer implique l'administration de protocoles de traitement complexes accompagnée d'une surveillance constante de la toxicité et de l'efficacité. De nombreuses variables (marqueurs pharmacogénétiques, mutations tumorales, caractéristiques cliniques, etc.) peuvent être impliquées dans le choix d'un protocole de traitement anticancéreux. Habituellement, chaque patient reçoit sa première dose de chimiothérapie selon un protocole en fonction de son poids, de sa taille, de son âge et de ses biomarqueurs. Au cours des cycles de traitement ultérieurs, des variables telles que la perte de poids, le score OMS, la ligne de traitement, la réponse au traitement, les polymorphismes pharmacogénétiques et les toxicités peuvent influencer la décision médicale de réduire la dose ou de choisir une autre ligne de traitement. Des études réalisées dans la vie réelle concordent avec le fait qu'en raison des toxicités attendues ou observées des médicaments, il existe de nombreux écarts aux protocoles, qu'il s'agisse des doses ou des délais entre les cycles.

Cependant, en vie réelle, les médecins ne peuvent compter que très partiellement sur des directives pour adapter leur décision médicale. En effet, les recommandations sont en général issues des essais cliniques. Or toutes les sous-populations de patients ne participant pas aux essais cliniques, une grande hétérogénéité est observée entre les données issues des essais cliniques et les données de routine. L'absence de recommandations précises en pratique en matière de réduction de dose pour certains sous-groupes de patients est donc due au grand nombre de variables influençant les toxicités et au grand nombre de paramètres intervenant dans le schéma posologique, à savoir les doses et les reports de cycles. L'influence de tous ces paramètres est difficile à modéliser dans la vie réelle. Fournir des lignes directrices ou des outils d'aide à la décision concernant la réduction de la posologie de la chimiothérapie pour différents sous-groupes de patients implique de s'attaquer à la grande dimensionnalité des données réelles et de modéliser la relation complexe entre les caractéristiques des patients, les biomarqueurs, y compris les marqueurs pharmacogénétiques, le schéma posologique et la toxicité/efficacité. Pour mieux comprendre cette relation complexe, la première étape consiste à modéliser la décision médicale de réduction de la dose.

La modélisation des décisions médicales peut être réalisée à l'aide d'approches fondées sur les données ou sur les modèles. Il existe des algorithmes explicites intégrant des capacités de raisonnement basées sur des moteurs d'inférence et des bases de connaissances, y compris des règles validées, des directives partagées et des ontologies. Cependant, la décision médicale est un processus assez complexe en pratique

clinique et toutes les décisions ne peuvent être modélisées à l'aide de modèles basés sur des données ou basées sur la connaissance.

Chaque jour, les médecins peuvent optimiser leur processus de prise de décision médicale en fonction des caractéristiques des patients et de leur expérience, car ils ne peuvent compter que très partiellement sur des recommandations. Par conséquent, certaines variables peuvent être prises en compte dans les soins de routine, bien qu'elles n'apparaissent pas dans les modèles basés sur la connaissance ou soient trop dispersées dans le jeu de données, de sorte que l'algorithme habituel ne peut pas les identifier comme étant importantes. En conséquence, grâce aux performances croissantes des techniques de *data mining* et d'apprentissage automatique, des algorithmes basés sur des analyses de données de grande dimension et de routine doivent être développés dans le secteur de la santé.

Par ailleurs, en oncologie, de nombreuses molécules et biomarqueurs innovants (génomiques, cliniques ou physiologiques) ont récemment été proposés, ce qui a permis de définir de nombreuses petites sous-populations de maladies supposées auparavant homogènes. Dans ces petites sous-populations, l'identification des causes de toxicité et les modifications du traitement dans les soins de santé quotidiens peuvent être assez compliquées. Avec de petits échantillons, les modèles multivariés ajustés peuvent avoir de grandes variances de paramètres estimés, limitant la fiabilité inférentielle. Les méthodes de sélection de variables dans de tels contextes peuvent atténuer cette difficulté en ne conservant que les covariables ayant les effets les plus importants. La plupart de ces méthodes sont fréquentistes et utilisent des critères de pénalisation, tels que les régressions *bridge*, LASSO (*Least Absolute Shrinkage and Selection Operator*), SCAD (*Smoothly Clipped Absolute Deviation*), *elastic net* ou OSCAR. (*Octagonal Shrinkage and Clustering Algorithm for Regression*) [102, 33, 31, 118, 13]. L'une des méthodes les plus populaires est le LASSO, qui effectue une sélection de variables en appliquant une pénalité L1 aux moindres carrés. Le LASSO estime certains coefficients à zéro tout en sélectionnant d'autres variables. Cependant, ces méthodes de sélection de variables ont été validées et associées à de bonnes propriétés de convergence pour de grands échantillons. Dans un contexte d'échantillons de petites tailles associés à de nombreuses variables, ces méthodes ne permettent souvent pas d'obtenir des résultats fiables.

La régression bayésienne est une approche alternative raisonnable lorsque l'on travaille avec de petits échantillons. Dans le cadre bayésien, la sélection de variables peut être effectuée en utilisant des distributions a priori de type "*spike and slab*" pour identifier et estimer les distributions a posteriori des paramètres de régression non nuls, tout en écartant les variables qui ne semblent pas prédictives de la variable dépendante [71]. Mitchell et Beauchamp ont considéré, comme distribution a priori pour chaque coefficient de régression, un mélange à deux composantes comprenant une masse ponctuelle en 0 (le "*spike*") et une distribution uniforme sur un intervalle fini (le "*slab*") [68]. Ces deux composantes quantifient les connaissances a priori qu'il existe un effet de la variable et, le cas échéant, sa magnitude. Cette méthode a été étendue de nombreuses manières, en se concentrant sur des formulations différentes des distributions a priori [38, 61, 55, 34, 117]. Une extension très utile est la méthode de *Stochastic Search Variable Selection* (SSVS), dans laquelle la distribution a priori de chaque coefficient est un mélange de deux distributions normales centrées en zéro mais avec des variances différentes, l'une petite et l'autre grande [38]. D'autres méthodes spécifient des distributions a priori continues qui se rapprochent de la forme "*spike and slab*" tout en réduisant les estimations à zéro [47, 73]. Park et Casella ont étendu le lasso fréquentiste en considérant une double distribution exponentielle pour les paramètres de régression [73]. Un autre avantage de l'inférence bayésienne est que de l'information externe peut être introduite dans les distributions a priori.

Dans cette thèse, nous proposons deux méthodes Bayésiennes de sélection de variables permettant de combiner les données expertes élicitées à l'analyse des données réelles. Notre application porte sur l'adaptation des doses répétées d'irinotécan dans le cancer métastatique colorectal en vie réelle.

## Données ayant motivé le développement de méthodes

Pour étudier la relation complexe entre caractéristiques des patients, schémas posologiques et toxicités, le développement des dossiers de santé électroniques et des entrepôts cliniques dans les hôpitaux fournit une source importante de données observationnelles. Ces entrepôts de données ont été créés pour faciliter la réutilisation des données de soins de santé aux fins de recherche [35, 99]. Nous avons extrait des dossiers de santé informatisés de l'entrepôt i2b2 de l'hôpital universitaire Georges Pompidou (HEGP) à Paris des données de patients traités pour un cancer colorectal métastatique avec une combinaison de médicaments incluant l'irinotécan. Le protocole comprend une dose théorique de 180 mg/m<sup>2</sup> et un cycle théorique de chimiothérapie de 14 jours. Tous les patients ayant été traités avec un autre protocole incluant de l'irinotécan ont été exclus. Un cycle a été défini comme les doses ayant été données sur une période allant de un à trois jours. Si un patient reçoit une dose plus de 28 jours après, nous avons considéré qu'il s'agissait du commencement d'un autre protocole. Les données incluent les variables de l'âge, du poids, de la bilirubine totale, du score OMS, de la ligne de traitement et les sept types de toxicité enregistrés sur chaque cycle de chimiothérapie. Le projet a été approuvé par le comité d'examen institutionnel HEGP, CERHUPO (CPP Ile de France II, numéro d'enregistrement IRB #00001072) sous la référence 2014-12-08.

## Modélisation de l'expertise : Principe de l'élicitation

Les poids de sévérité reflètent l'importance qu'un médecin place dans chaque valeur de chaque variable dans sa décision médicale [6]. Dans notre situation, les oncologues nous ont aidé à construire un questionnaire en spécifiant les variables et les seuils qu'ils utilisent dans leur prise de décision. Puis, six médecins experts ont complété le questionnaire en ligne séparément. Chacun des médecins a spécifié un poids de sévérité numérique  $w_l^i$  pour chaque niveau de chaque variable et chaque grade chaque type de toxicité  $l$ ,  $l \in \{1, \dots, L\}$ , dans l'intervalle  $[0, 100]$ , avec 0 et 100 qui correspondent respectivement à une absence et à la plus importante gravité.

## Stratégies développées pour combiner les données expertes et les données observées

### Premier modèle développé : Le modèle *Weights-Based SSVS* (WBS)

#### Méthode

Afin d'identifier les variables ayant une influence sur l'ajustement de dose, la relation entre les réductions de dose répétées en fonction des caractéristiques des patients et des toxicités mises à jour à chaque cycle, est modélisée par un modèle linéaire à effets mixtes dans le premier article. Soit  $d_{i,k}$  la dose donnée au patient  $i$  sur le cycle  $k$  et  $\mathbf{z}_{i,k}$  le vecteur des variables du patient à chaque visite, pour  $i \in \{1, \dots, n\}$  et  $k \in \{1, \dots, K\}$ . Le modèle est donné par l'équation :

$$d_{i,k} = \theta_0 + \boldsymbol{\theta}^T \mathbf{z}_{i,k} + \gamma_i + \epsilon_{i,k} \quad (5.1)$$

où  $\boldsymbol{\theta}$  est un vecteur de paramètres de longueur  $L$ ;  $\gamma_1, \dots, \gamma_n$  sont des effets patients aléatoires iid  $\mathcal{N}(0, \sigma_\gamma^2)$ ; et  $\epsilon_{i,k} \sim \mathcal{N}(0, \sigma_\epsilon^2)$  est le terme d'erreur. Avec cette formulation,  $\mathbf{z}_{i,k}$  inclut les indicatrices des grades des toxicités qui ont eu lieu sur le cycle  $k - 1$  et chacune des variables initiales mises à jour. Notons que les toxicités ne sont pas les variables dépendantes du modèle de régression, mais les variables explicatives.

Notre objectif est de déterminer un sous-ensemble,  $\mathcal{I}$ , du vecteur  $\mathbf{z}$  décrivant correctement les données.



Tableau 5.1 – Valeurs de  $(a, b)$  intervenant dans la loi beta centrée sur les poids élicités/100.

Poids élicités	0	10	20	30	40	50	60	70	80	90	100
$a$	1.41	1.41	1.41	2.88	4.43	5.78	6.65	6.72	5.64	5.64	5.64
$b$	5.64	5.64	5.64	6.72	6.65	5.78	4.43	2.88	1.41	1.41	1.41

Pour la variable d'indice  $l \in \{1, \dots, L\}$ , notons  $I_l$  l'indicatrice de sélection de la variable prenant la valeur

$$I_l = \begin{cases} 1 & \text{si } z_{..l} \text{ est dans le modèle} \\ 0 & \text{sinon,} \end{cases}$$

où  $z_{..l}$  fait référence à la  $l$ ème variable par commodité. Le modèle SSVS suppose que la distribution a priori de chaque paramètre du modèle (5.1) est un mélange de deux distributions normales, toutes les deux centrées en zéro mais avec des variances différentes [38] :

$$\theta_l | I_l \sim (1 - I_l)\mathcal{N}(0, \tau_l^2) + I_l\mathcal{N}(0, g_l\tau_l^2) \quad (5.2)$$

où  $P(I_l = 1) = p_l$ , c'est à dire  $I_l \sim \mathcal{B}(p_l)$ . Dans l'équation 5.2, le premier terme (le "spike") correspond aux variables  $z_{..l}$  ( $I_l = 0$ ) non sélectionnées, avec une densité centrée en 0 ayant une petite variance. Le second terme correspond aux variables sélectionnées  $z_{..l}$ , où  $I_l = 1$ , et a donc une large variance. Dans le contexte usuel de la méthode SSVS,  $p_l$  est fixé égal à 0,5. Dans cette première méthode, nous utilisons les poids élicités par les oncologues pour construire une distribution a priori pour  $I_l$ , de telle sorte que  $I_l$  ait plus de probabilité d'être égale à 1 quand les poids élicités sont importants, et plus de probabilité d'être égale à 0 autrement. Plutôt que de fixer  $p_l$ , nous supposons donc qu'il suit une distribution a priori Beta :

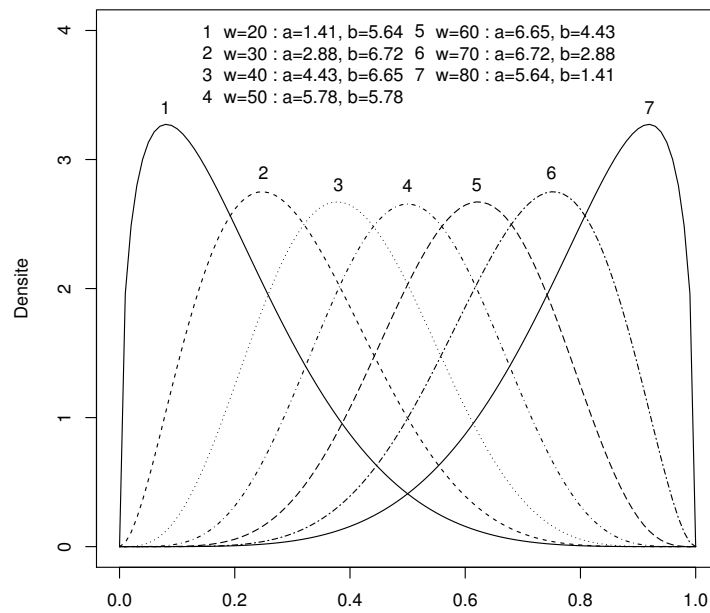
$$p_l \sim \text{Beta}(a_l, b_l) \quad (5.3)$$

Par conséquent,  $I_l | p_l$  suit une distribution beta-binomiale de paramètres  $(a_l, b_l)$ . La principale innovation de notre méthode, appelée *Weights-Based SSVS* (WBS), est l'introduction de cette équation 5.3 et le calcul des hyperparamètres  $a_l$  et  $b_l$  à partir des poids  $w'_l$  élicités par les cliniciens. En résumé, chaque clinicien commence par spécifier séparément un vecteur de poids élicités numérique  $\mathbf{w}'$  associés à chaque grade de chaque type de toxicité et à chaque niveau de chaque variable. Puis,  $a_l$  et  $b_l$  sont déterminés en résolvant les équations :

$$\begin{cases} \mathbb{E}(p_l) = \frac{a_l}{a_l + b_l} = \frac{w'_l}{W_{max}} \\ \mathbb{V}(p_l) = \frac{a_l b_l}{(a_l + b_l)^2 (a_l + b_l + 1)} = S. \end{cases}$$

La moyenne de  $p_l$  est donc  $w'_l/W_{max}$ , avec  $S$  un paramètre de réglage prenant en compte l'incertitude, choisi suite à une analyse de sensibilité. Notons qu'aucune procédure d'inférence n'est nécessaire pour éliciter les poids. Pour implémenter la méthode WBS, des seuils minimal et maximal sont fixés à 20 et 80 respectivement, afin d'éviter des distributions a priori trop informatives. Les valeurs de  $a$  et  $b$  obtenues à partir des poids élicités utilisés dans la méthode WBS sont présentées dans le Tableau 5.1 et au Graphique 5.1. Nous avons choisi  $W_{max} = 100$  car l'échelle  $[0, 100]$  est bien connue et facilement compréhensible des médecins, et nous avons fixé  $S = 0,02$  suite à une analyse de sensibilité préliminaire.  $S$  quantifie la variabilité que l'on souhaite ajouter à la distribution a priori. La méthode WBS utilisant les poids élicités par le même médecin est noté WBS( $m$ ).

Afin de combiner l'expertise des différents médecins, nous avons utilisé deux méthodes : (1) une loi de mélange obtenue en pondérant chacune des distributions a priori élicitées par les médecins par le



Graphique 5.1 – Densité de la distribution beta de  $p$  de paramètres  $a$  et  $b$  et de moyenne  $w/100$ .

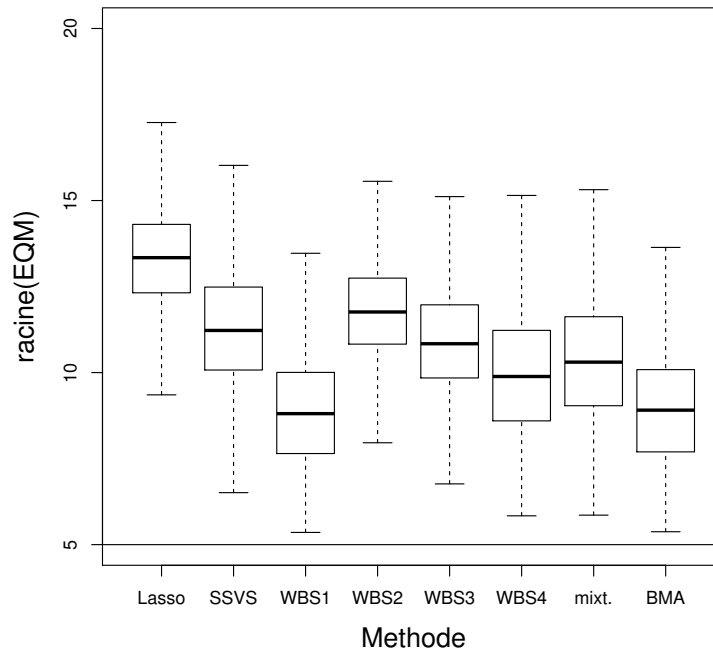
nombre de médecins, et (2) la méthode du *Bayesian Model Averaging* (BMA) sur les modèles obtenus à partir des poids élicités pour chacun des médecins.

## Résultats

Les poids élicités varient beaucoup selon les cliniciens, ce qui peut souligner une importante variabilité des pratiques professionnelles. Cela peut aussi indiquer que les cliniciens ne perçoivent pas correctement leur réel processus de décision pour l’ajustement de dose. En effet, beaucoup de gens sont influencés par des facteurs cognitifs, psychologiques ou émotionnels qui les empêchent parfois de choisir l’option la plus rationnelle possible. Par exemple, en pratique clinique, si un médecin rend visite à un patient avec de forts effets indésirables un jour, ses émotions et son expérience peuvent le mener à réduire la dose du prochain patient, même si le bon sens serait de maintenir la dose actuelle.

Les performances du modèle WBS ont été comparées à celles des modèles LASSO et SSVS par l’intermédiaire d’une étude de simulation. Pour quatre scénarios différents, 1000 jeux de données ont été simulés. Chaque jeu de données consistait en un échantillon de patients traités par irinotécan pour un cancer colorectal métastatique et était simulé selon les distributions de notre jeu de données observé. Le Graphique 5.2 présente la distribution de la racine de l’erreur quadratique moyenne pour chacune des méthodes (LASSO, SSVS, modèle WBS pour chacun des cliniciens, avec modèle de mélange et avec BMA) pour un scénario pour un échantillon de  $n = 10$  patients et  $K = 3$  cycles. Le Graphique 5.3 donne le pourcentage de sélection de chaque variable par chacune des méthodes.

Notre méthode WBS montre de meilleures performances que la méthode SSVS pour les différents critères utilisés : biais, erreur quadratique moyenne, taux de faux positifs et de faux négatifs. Les performances de cette méthode dépendent des poids élicités par les médecins pour chaque variable et en particulier de leur somme qui doit donc être calibrée de manière à sélectionner le nombre souhaité de variables. Les performances du LASSO sont faibles pour de petits échantillons, confirmant l’attrait de l’inférence Bayésienne dans ce cadre.



Graphique 5.2 – Boxplot de la racine de l’erreur quadratique moyenne pour les méthodes LASSO, SSVS et WBS pour un scénario avec  $n = 10$  patients et  $K = 3$  cycles.

La méthode WBS a fait l’objet d’une publication dans *Statistics in Medicine* en 2019. J’ai également présenté ce travail en 2018 lors de communications orales à la conférence francophone d’Épidémiologie CLINique (EPICLIN) à Nice et à la *Joint International Society for Clinical Biostatistics and Australian Statistical Conference* (ISCB ASC) à Melbourne (Australie), et sous forme d’un poster à l’*International Society for Bayesian Analysis* (ISBA) à Édinburgh (Écosse).

## Deuxième modèle développé : Modèle SSVS couplé au *Power Prior* (PP)

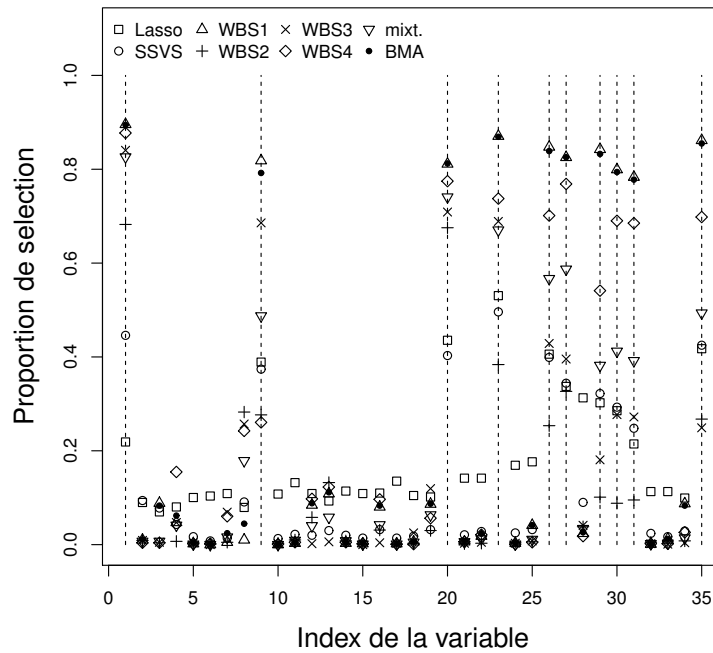
### Méthode

Dans le second article, nous utilisons un modèle de régression logistique à effets mixtes où la réduction de dose pour un patient  $i$  au cycle  $k$  est une variable binaire  $r_{i,k}$ , et le vecteur de variables  $\mathbf{z}_{i,k}$  inclut les caractéristiques des patients et les toxicités, mises à jour sur le cycle précédent, comme précédemment :

$$\begin{aligned} r_{i,k} &\sim \mathcal{B}(p_{i,k}^r) \\ \text{logit}(p_{i,k}^r) &= \theta_0 + \boldsymbol{\theta}^T \mathbf{z}_{i,k} + \gamma_i \\ \gamma_i &\stackrel{iid}{\sim} \mathcal{N}(0, \sigma_\gamma^2) \end{aligned}$$

où  $(\theta_0, \boldsymbol{\theta})$  est le vecteur des effets fixes de longueur  $L$ ,  $\gamma_i$  correspond aux effets patients aléatoires et  $\mathcal{B}(p^r)$  représente une distribution de Bernoulli dont la probabilité de succès est égale à  $p^r$ . La sélection de variables est encore réalisée grâce au modèle SSVS mais une distribution normale tronquée est utilisée de manière à restreindre les paramètres à des valeurs négatives, de telle sorte que le modèle ne puisse considérer que des réductions de dose :

$$\theta_l | I_l \sim (1 - I_l)TN(0, \tau^2) + I_l TN(0, g\tau^2) \quad (5.4)$$



Graphique 5.3 – Pourcentage de sélection de chaque variable par chacune des méthodes LASSO, SSVS et WBS pour un scénario avec  $n = 10$  patients et  $K = 3$  cycles. Les variables utilisées pour simuler les données sont identifiées par des lignes en pointillé.

où  $T\mathcal{N}$  est une distribution normale tronquée sur  $(-\infty, 0]$ . En pratique, les doses spécifiées par les protocoles ne peuvent que décroître et non augmenter. Il n’y a donc pas d’augmentations de doses dans nos données.

Pour combiner les poids élicités experts aux données observées, nous utilisons la méthode du *Power Prior* (PP) [50]. La méthode du PP permet d’intégrer les informations contenues dans des données historiques dans une analyse bayésienne. Dans la méthode que nous avons proposée, les données historiques sont remplacées par des données expertes simulées générées à l’aide des patients dont les données sont incomplètes. Pour ces patients, nous avons d’abord (1) imputé les valeurs manquantes en fonction de la distribution des variables dans l’ensemble de données initial et (2) des réductions de dose  $r_0$  ont été générées en fonction des poids élicités.

Deux approches alternatives ont été utilisées pour générer les réductions de doses  $r_0$  en fonction des poids élicités par les experts :

— Approche 1 :

$$r_{0,i',k'}^{(1)} = \begin{cases} 1 & \text{si } \mathbf{w}^{*T} \mathbf{z}_{0,i',k'} > Q_0 \\ 0 & \text{sinon} \end{cases} \quad (5.5)$$

— Approche 2 :

$$\begin{aligned} r_{0,i',k'}^{(2)} &\sim \mathcal{B}(p_{0,i',k'}^r) \\ \text{logit}(p_{0,i',k'}^r) &= -Q_0 + \mathbf{w}^{*T} \mathbf{z}_{0,i',k'} \end{aligned} \quad (5.6)$$

Dans les deux approches,  $Q_0$  est un seuil défini pour obtenir une proportion de réductions de dose  $P_0$ , où  $Q_0$  est le quantile correspondant à la probabilité  $1 - P_0$  de l’échantillon composé de  $(\mathbf{w}^{*T} \mathbf{z}_{0,i',k'})_{i',k'}$ . Les équations 5.5 et 5.6 montrent comment les poids obtenus ont été utilisés comme paramètres de covariables dans le modèle de simulation de données expertes. De plus, le jeu de données simulé dépend

de la valeur  $P_0$  choisie en fonction d'une analyse de sensibilité. La seconde approche introduit une incertitude dans la réduction de la dose et rend le processus de prise de décision plus similaire à celui des cliniciens en pratique clinique en écartant parfois des variables importantes.

Les données simulées sont ensuite combinées aux données observées en utilisant le PP. La distribution a priori est définie comme le produit de la vraisemblance pondérée de  $\theta$  conditionnellement aux données expertes simulées et d'une distribution a priori non-informative :

$$\pi(\theta|\mathbf{D}_0, a_0) \propto L(\theta|\mathbf{D}_0)^{a_0} \pi_0(\theta)$$

où  $\mathbf{D}_0 = (n_0, \mathbf{r}_0, \mathbf{z}_0)$  correspond aux données expertes simulées,  $L(\theta|\mathbf{D}_0)$  est la vraisemblance de  $\theta$  conditionnellement aux données expertes simulées,  $0 \leq a_0 \leq 1$  est le paramètre d'échelle qui contrôle le poids donné aux données expertes simulées par rapport aux données observées et  $\pi_0(\theta)$  est une distribution a priori non informative pour  $\theta$ .

La distribution a posteriori de  $\theta$  est

$$\begin{aligned} \pi(\theta|\mathbf{D}, \mathbf{D}_0, a_0) &\propto L(\theta|\mathbf{D})\pi(\theta|\mathbf{D}_0, a_0) \\ &\propto L(\theta|\mathbf{D})L(\theta|\mathbf{D}_0)^{a_0}\pi_0(\theta) \end{aligned}$$

où  $\mathbf{D} = (n, \mathbf{r}, \mathbf{z})$  correspond aux données observées,  $L(\theta|\mathbf{D})$  est la vraisemblance de  $\theta$  conditionnellement aux données observées.

Enfin, de manière à combiner les poids élicités par les différents experts, la méthode BMA a de nouveau été utilisée.

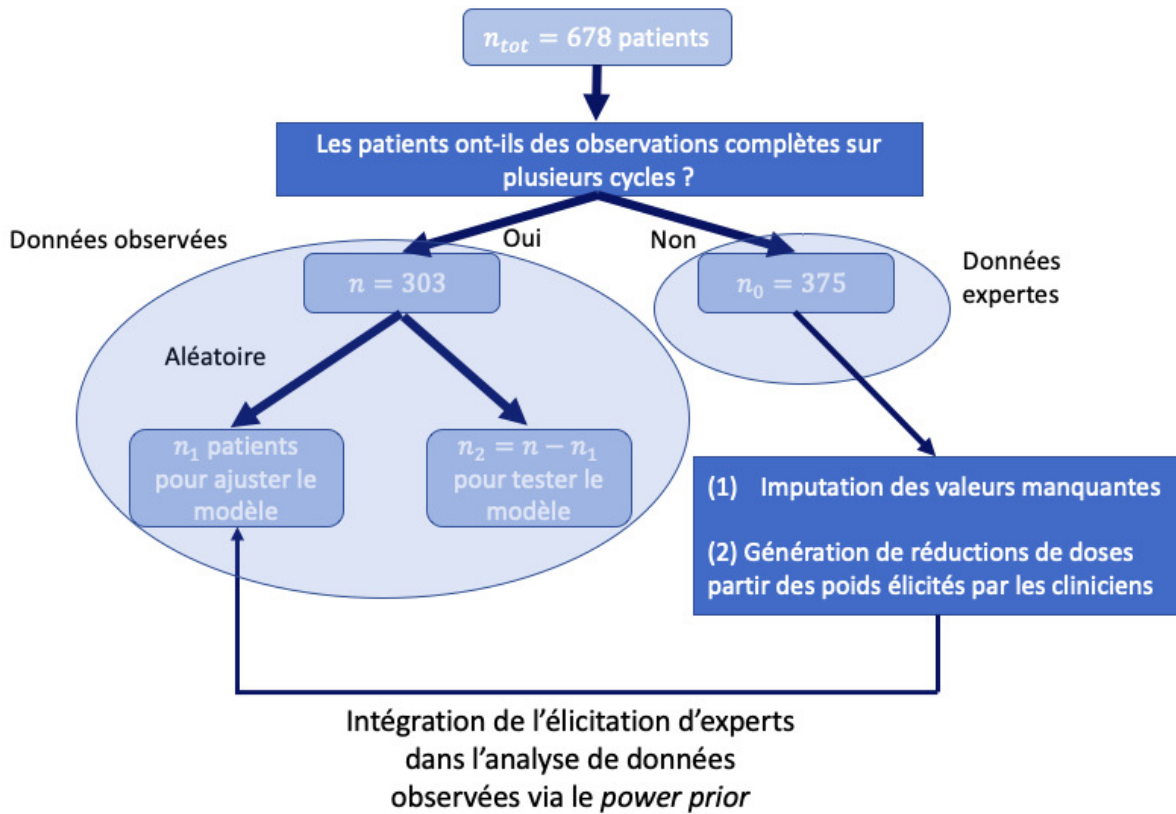
## Résultats

Parmi les 678 patients inclus dans l'étude,  $n = 303$  ont des données complètes sur plusieurs cycles et sont utilisés comme données observées à analyser. Dix échantillons d'entraînement de  $n_1 = 212$  patients et dix échantillons de validations de 91 patients ont été construits pour évaluer les performances prédictives. Les 375 patients avec des données incomplètes ont été utilisés pour construire le jeu de données expertes en (1) imputant leurs valeurs manquantes et en (2) générant les réductions de doses à partir des poids élicités par les médecins. Un *flowchart* récapitulatif l'utilisant des données est fournie au Graphique 5.4.

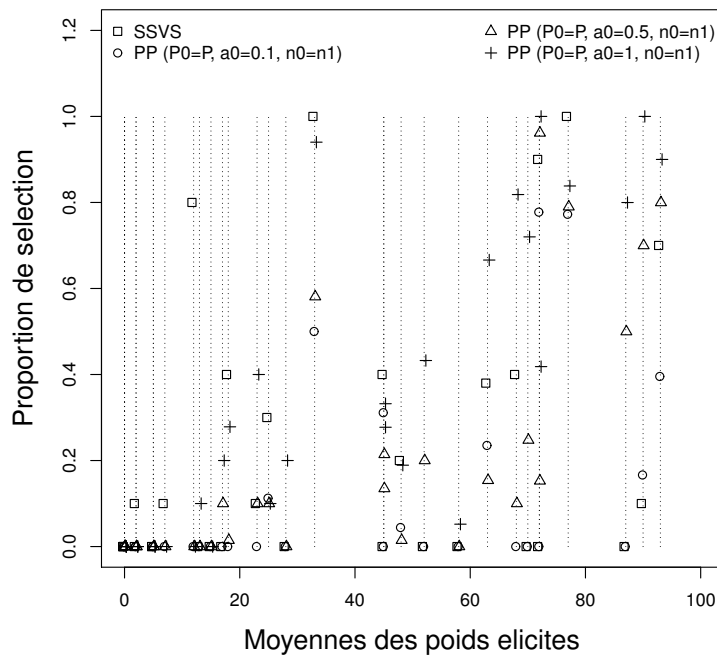
Nous avons étudié les performances du modèle SSVS couplé au PP pour différentes valeurs du paramètre  $a_0$ , différentes tailles d'échantillons simulés  $n_0$  et différentes fréquences de réductions de dose simulées  $P_0$ . La sélection de variables ne change pas significativement avec  $n_0$  et  $P_0$ . Le Graphique 5.5 donne la proportion de sélection des variables en fonction des poids moyens élicités par les experts pour chacun des modèles représentés par un symbole différent, pour un effectif de patients simulés et une proportion de réductions de doses simulées égaux à ceux des données observées et pour différentes valeurs de  $a_0$ . Chaque ligne en pointillées est associée à une variable différente.

Seules cinq variables sont sélectionnées par le modèle SSVS usuel plus d'une fois sur deux. Parmi elles, trois variables sont liées à un poids moyen élicité supérieur à 50. En utilisant l'approche du PP, les variables sélectionnées par la méthode SSVS sont encore sélectionnées, à l'exception de l'anémie de grade 2, associée à un poids moyen élicité de 12. En utilisant la même quantité de données observées et de données expertes simulées, le modèle sélectionne des variables avec des poids élicités importants et peu d'observations, telles que le score OMS de grade 3 (11 observations) et les vomissements de grade 3 (2 observations). Il sélectionne aussi le score OMS de grade 4, l'asthénie de grade 3 et l'âge. On peut noter que la performance prédictive est modérée avec un AUC proche de 0,7 pour les différents modèles.

Notre application montre donc que cette méthode sélectionne des variables rares qui ne pourraient pas être sélectionnées sur la base des seules données patients. La variabilité des pratiques cliniques



Graphique 5.4 – Flowchart de l'utilisation des patients pour le modèle SSVS couplé au *power prior*.



Graphique 5.5 – Pourcentage de sélection de chaque variable par le modèle SSVS couplé au *power prior* pour l'approche 1 et différentes valeurs de  $a_0$  ( $a_0 \in \{0, 1; 0, 5; 1\}$ ).

pourraient expliquer la performance prédictive modérée des modèles.

Ce second travail a été publié dans *Statistical Methods in Medical Research* en 2019. Je l'ai présenté en communication orale à EPICLIN à Toulouse cette même année. Je le présenterai également à l'oral au colloque de Données de Santé en Vie Réelle (DSVR) à Paris et à l'*International Symposium of Biopharmaceutical Statistics* à Kyoto (Japon), et sous forme de poster à l'ISCB à Louvain (Belgique) cette année.

---

# Conclusion

Aujourd'hui, aucune méthode automatique ne permet de sélectionner des variables d'intérêt en combinant expertise et données de vie réelle. Les méthodes Bayésiennes développées dans cette thèse proposent de combiner l'information experte aux données observées afin de sélectionner l'ensemble optimal des variables importantes intervenant dans une prise de décision. Plus précisément, l'information préalablement élicitée par les experts est introduite sous forme de poids numériques dans les paramètres des distributions a priori.

Il est souvent reproché aux analyses en vie réelle d'inclure toutes les variables disponibles dans la base de données mais pas celles que les médecins utilisent pour prendre leurs décisions en pratique. En intégrant l'expertise clinique dans l'analyse des données observées, nous nous assurons qu'aucune variable importante pour le clinicien n'est omise. Par ailleurs, les méthodes développées ne sont pas fondées sur un consensus entre experts et peuvent donc s'appuyer sur différentes expertises.

Dans cette thèse, nous ne nous sommes pas intéressés à la performance prédictive des modèles mais bien à relever toutes les variables qui pouvaient s'avérer pertinentes dans la prise de décision. Dans les analyses ne s'appuyant que sur les données observées, certaines variables peu intéressantes pour les experts peuvent ressortir. En présence de variables relativement corrélées, les stratégies exposées ici peuvent permettre de conserver en priorité les variables réellement utilisées dans la pratique clinique. La méthodologie présentée pourrait ainsi être utilisée pour construire des outils d'aide à la décision en permettant de sélectionner les variables impliquées dans le processus de prise de décision. Intégrer l'information experte dans l'analyse des données observées peut autoriser un outil d'aide à la décision à être plus largement diffusé en évitant l'aspect "boîte noire" et en lui permettant d'être plus facilement accepté par les médecins.

Une limite de la première méthode (méthode WBS du chapitre 4) est qu'elle ne laisse pas suffisamment la possibilité de moduler l'expertise pour des échantillons importants. La seconde méthode (cf. chapitre 5) pallie à cette problématique. Les deux articles présentés diffèrent ainsi par plusieurs aspects : (1) D'abord, l'expertise clinique n'est pas introduite dans le modèle de la même manière. En effet, dans la méthode WBS, nous utilisons les poids de pertinence clinique pour calculer les hyper-paramètres des distributions a priori des indicatrices de sélection de variable intervenant dans la méthode SSVS. Dans la méthode utilisant le PP, l'expertise clinique est utilisée pour générer des données expertes qui sont ensuite jointes aux données observées grâce au PP (toujours dans le cadre du modèle de sélection de variables SSVS). Le cadre est donc le même mais l'information est introduite de deux manières différentes. (2) Puis, l'ensemble des données réelles considéré dans le second projet est plus important que celui du premier projet (303 patients contre 10 à 20 patients). Par conséquent, dans l'article du chapitre 4, nous utilisons un modèle de régression linéaire à effets mixtes au lieu du modèle logistique à effets mixtes de l'article du chapitre 5. Le modèle linéaire du premier projet nous permettait de conserver la totalité de l'information présente dans les données de notre petit échantillon. En considérant un échantillon plus grand, nous pouvons utiliser un modèle logistique qui, en binarisant la réduction de doses selon un seuil fixé, nous permet de mieux refléter l'aspect modal de la réduction de doses qui intervient dans l'esprit des médecins (par exemple, réductions de 25% ou 50 % par rapport à la dose



initiale). Ainsi, la variable dépendante est l'indicatrice binaire de réduction de dose plutôt qu'une dose continue comme c'était le cas dans le premier article.

Les méthodes proposées ont été appliquées à des données de chimiothérapie faisant intervenir des ajustements de doses d'irinotécan répétés au cours des cycles de traitement pour des patients atteints de cancer colorectal métastatique et sujets à des effets indésirables. Nous avons montré qu'elles permettaient de sélectionner des variables rares qui pouvaient être omises par les méthodes de sélection de variables usuelles à cause d'un manque de puissance, et de rejeter des variables que les experts estiment peu pertinentes.

Les méthodes développées pourraient être facilement appliquées à d'autres modèles et d'autres situations pour lesquelles on souhaite prendre en compte l'opinion d'experts dans la sélection de variables, comme par exemple les modèles de survie ou des données non-longitudinales.

---

# Perspectives

Plusieurs perspectives d'extension de ce travail sont actuellement étudiées en collaboration avec Antoine Chambaz, professeur à l'Université Paris Descartes et membre de son laboratoire de mathématiques appliquées (MAP5), Agathe Guilloux, professeur à l'Université d'Évry Val d'Essone et membre du Laboratoire de Mathématiques et Modélisation d'Évry (LaMME) et Satoshi Morita, professeur au département de sciences des données, de l'Institut pour l'avancement des sciences cliniques et translationnelles de l'hôpital universitaire de Kyoto (Japon).

## Contexte clinique et objectif

Dans ce nouveau projet, nous conservons le cas d'usage de la relation entre l'administration des doses de chimiothérapie, les caractéristiques des patients et les toxicités pour les patients atteints d'un cancer colorectal métastatique.

Dans ce contexte, en fonction des effets indésirables subis par le patient sur un cycle donné, le médecin peut choisir de réduire la dose sur le cycle suivant mais il peut aussi préférer allonger l'intervalle de temps avec le cycle suivant, c'est à dire l'intervalle de temps avec la dose suivante, sans pour autant réduire cette nouvelle dose. Les intervalles de temps entre deux doses peuvent donc varier entre les patients mais aussi pour un même patient. Cette possibilité n'a, jusqu'à présent, pas été prise en compte dans nos modèles d'analyse et nos plans de simulation, étant donné que la variable du temps n'est pas introduite. D'autre part, les oncologues considèrent tous les traitements et effets indésirables du patient sur tous les cycles précédents pour choisir la dose et le prochain jour de visite. Dans nos modèles, nous n'avons pris en compte que les variables actualisés sur le dernier cycle. Enfin, les toxicités dépendent directement des doses précédentes et des temps de visite. Il y a donc un lien cause-effet double qui n'est pas pris en compte.

Par ailleurs, en pratique clinique, un protocole peut être arrêté pour plusieurs raisons : la progression de la maladie, une toxicité (aiguë ou cumulée) et le succès du protocole. Tous les patients atteints d'un cancer métastatique connaîtront une progression de la maladie mais la survie sans progression (*Progression Free Survival* - PFS) pour un protocole spécifique peut être censuré. Si trop de toxicités dues au traitement sont observées, les médecins peuvent décider d'arrêter le protocole actuel mais aussi de réduire les doses administrées. De plus, une bonne réponse thérapeutique et/ou une diminution des marqueurs après plusieurs cycles sous le même protocole peuvent également amener les médecins à opter pour une pause thérapeutique synonyme de succès. Néanmoins, l'arrêt du protocole dans une telle situation dépend à la fois des médecins et des patients. Si un patient reçoit de nombreux cycles du même protocole (pendant plus d'un an par exemple), cela peut être considéré comme une réussite du protocole à chaque cycle jusqu'à ce que sa progression soit interrompue. Dans mon travail de thèse, la modélisation était simplifiée et ne prenait en compte ni l'arrêt d'un protocole et les raisons de cet arrêt, ni les pauses thérapeutiques, ni les toxicités cumulées, ni la progression de la maladie. Ce nouveau projet implique donc une structure de données plus riches et comprenant une variable de temps.

Dans le contexte de la chimiothérapie, la structure de données temporelle pourrait donc être décomposée à chaque temps  $t$  en :

- Un vecteur de covariables  $W(t)$  comprenant les caractéristiques du patient à un temps donné, comme par exemple une mesure de la concentration de la bilirubine et le score OMS ;
- Un traitement  $A(t)$  indiquant la dose attribuée à un temps donné ;
- Une variable dépendante consistant en sept types de toxicité gradées  $T(t)$  pouvant être subies par le patient en raison du traitement administré au cours du temps et de la progression de la maladie  $Y(t)$ .

La question centrale concernant la stratégie de décision clinique consiste à déchiffrer les rôles respectifs ou les effets de causalité des caractéristiques du patient et du régime de dose sur les toxicités et la progression de la maladie. En l'absence de recommandation personnalisée, il existe aujourd'hui une grande variabilité de pratique clinique que nous voulons réduire en développant des prototypes d'outils d'aide à la décision cliniques suggérant des propositions de régimes de doses adaptatifs. En effet, du point de vue du patient, une telle variabilité implique l'hétérogénéité des traitements d'un clinicien à l'autre et, dans certains cas, un traitement non optimal, ce qui entraîne des chances inégales de succès du traitement.

L'objectif de ce projet est donc d'identifier les règles de traitement optimales pour des chimiothérapies à l'aide de données observationnelles provenant des dossiers de santé électroniques, c'est-à-dire des règles permettant d'augmenter la proportion de patients présentant une progression à long terme, tout en maintenant les toxicités à des grades faibles. Il s'agit d'une étape essentielle vers des stratégies de traitement personnalisées. Plusieurs problèmes doivent être étudiés : l'inférence des paramètres de causalité à la fois pour la toxicité et la progression de la maladie, la censure de la variable dépendante et des intervalles de temps irréguliers entre les cycles. Nous profiterons du développement très récent en intelligence artificielle, en particulier en apprentissage automatique, pour répondre à ces problématiques.

## Contexte scientifique

Le projet s'appuiera notamment sur les développements théoriques récents permettant d'inférer des paramètres de causalité à partir de séries temporelles structurées et des données longitudinales [103, 104, 75, 87, 82].

Nous définissons une série temporelle structurée comme une série temporelle où sont observés à chaque temps, dans l'ordre chronologique, un vecteur de covariables  $W(t)$ , un traitement  $A(t)$  et un résultat  $Y(t)$ . Nous supposons que la distribution de la probabilité conditionnelle de cette structure de données temporelles, sachant le passé, en dépend par le biais d'une mesure de synthèse dimensionnelle fixe (dans le temps), et que cette distribution conditionnelle est décrite par un mécanisme fixe (dans le temps) qui est connu pour être un élément d'un espace modèle. En l'absence de connaissances élaborées et spécifiques au sujet, l'espace modèle pourrait être non spécifié, c'est-à-dire être le modèle non paramétrique de tous les mécanismes compatibles avec la structure de données, ce qui permettrait d'obtenir plus facilement une inférence robuste.

Les développements récents en matière d'apprentissage ciblé pour les données longitudinales proposent de prendre en compte le fait que toutes les variables observées peuvent être censurées. Ces modèles ont été développés dans le contexte des données de dossiers de santé électroniques, comme c'est le cas dans notre projet.

L'identifiabilité des quantités causales à partir de la distribution de données observée est établie sous des conditions qui sont typiques de la littérature sur l'analyse causale [74, 105, 106].

Les méthodologies existantes pour les séries temporelles structurées et les données longitudinales permettent de modéliser l'acquisition de données longitudinales, ou le mécanisme de censure, spécifique aux dossiers médicaux électroniques. Cependant, aucun modèle mathématique englobant les deux aspects de ces données n'a été proposé, à notre connaissance.

## Méthodes

Pour atteindre notre objectif, nous procéderons en quatre étapes.

1. Nous voudrions tout d'abord proposer un modèle mathématique avec une structure de dépendance complète, prenant en compte également tous les aspects spécifiques des données longitudinales des dossiers de santé électroniques, en particulier le mécanisme de censure. Ce modèle servira également à l'inférence ;
2. Sur cette base, nous comptons proposer un algorithme de simulation permettant de déduire des schémas de simulation jugés réalistes par les experts médicaux ;
3. Nous développerions ensuite une implémentation efficace de notre stratégie d'inférence ;
4. Enfin, nous souhaiterions développer un prototype d'outil d'aide à la décision clinique, entraîné à partir des données de l'entrepôt de données i2b2 de l'HEGP.

Les paragraphes suivants précisent quelques détails méthodologiques.

### Modèle mathématique

Premièrement, le développement de notre modèle mathématique inclura la dépendance présentée par des séries temporelles structurées [103, 104], mais également le mécanisme de censure de la même manière que cela a déjà été fait dans les modèles causaux pour les données longitudinales dans [75, 87, 82]. La structure de dépendance entre la toxicité et la progression de la maladie attirera particulièrement notre attention.

### Simulation

Nous sommes actuellement en train de développer un package sous  $R$  permettant une simulation réaliste des séries temporelles intervenant dans le contexte du cancer colorectal. La fonction principale du package peut être ajustée avec précision grâce au choix de dizaines de paramètres, à valeurs réelles ou sous forme de fonctions (donc à la fois de dimensions finies et infinies).

Nous voudrions étendre ce travail en développant une procédure permettant d'effectuer le réglage des paramètres de manière automatique en nous appuyant sur des réseaux antagonistes génératifs [40]. L'idée peut être résumée comme suit. Deux réseaux de neurones sont confrontés. L'un, appelé générateur, génère des séries temporelles synthétiques basées sur la fonction principale du package. L'autre, appelé discriminateur, évalue l'authenticité de ces séries temporelles en les comparant aux séries temporelles observées dans la vie réelle. Le générateur et le discriminateur entrent en compétition en tentant d'optimiser des fonctions objectives opposées. Lorsque le discriminateur modifie son comportement, le générateur fait de même, et inversement. Ils améliorent ainsi, de manière séquentielle, leurs performances. Finalement, le générateur peut simuler des séries temporelles qui semblent réalistes, même pour des experts médicaux.

### Inférence

Les quantités causales pourraient consister en une mesure de l'effet causal d'une politique de traitement donnée, c'est-à-dire une mesure de ce qui aurait été le résultat (contrefactuel) du traitement à un horizon donné si (contrairement aux faits) les doses avaient été déterminées de manière séquentielle sur la base de la politique de traitement étudiée. Pour atteindre notre objectif, une stratégie inférentielle en deux étapes pourrait être particulièrement adaptée. La stratégie consisterait, d'une part, à construire des estimateurs pertinents des caractéristiques complexes de l'expérience en population, nécessaires pour déduire les quantités causales d'intérêt, et, d'autre part, à adapter ces caractéristiques estimées à un seul patient en reprenant la procédure d'estimation fondée sur la série temporelle de ce patient. L'inférence serait donc totalement spécifique au patient. Dans un premier temps, nous proposons d'élargir la

stratégie d'estimation mise au point dans [106] à notre contexte de données longitudinales sujettes à la censure des dossiers de santé électroniques. Une autre extension méthodologique consisterait à proposer des modifications des modèles pour tenir compte de l'irrégularité des temps d'observation. Dans la deuxième étape, nous visons à adapter ces caractéristiques estimées à un seul patient en reprenant la procédure d'estimation basée sur la série temporelle de ce patient.

### **Prototype d'un outil d'aide à la décision clinique**

À partir des études de simulation, nous souhaitons mettre en évidence des situations pour lesquelles nous bénéficions d'une quantité de données suffisante pour déduire une décision clinique adaptative en matière de réduction de dose. Nous implémenterons ensuite la procédure d'estimation dans un outil d'aide à la décision connecté au dossier médical électronique. À l'aide de ce prototype, nous pourrions discuter avec un expert du domaine (un oncologue) de la pertinence de la réduction de dose estimée pour une dizaine de patients.

Cette méthodologie pourrait être testée sur les données Françaises et Japonaises puis mise à disposition de la communauté scientifique souhaitant travailler sur les trajectoires de doses. En effet, bien que tout traitement puisse être administré selon différents régimes de doses, il n'existe actuellement pas de méthodes de simulation satisfaisantes de tels processus dans la littérature.

À termes, ce travail pourrait permettre de proposer une méthodologie pour estimer le meilleur régime de dose selon les caractéristiques du patient et ses antécédents de traitement et donc de proposer un prototype d'outil d'aide à la décision clinique.

— Annexe A —

# Compte-rendu de CHIMIO-ONCODIG de DxCare

**OBSERVATION MEDICALE - Compte-rendu de CHIMIO-ONCODIG**

**Cure de Chimiothérapie** | **INTERCURE (Toxicités observées après la précédente cure)** | **EXAMEN CLINIQUE**

Date:

**Diagnostique**

Cancer:

Médecin(s) référent(s):

Principaux antécédents:

Autre pathologie:

Traitement habituel:

N° Cure de Chimiothérapie:

Type de traitement:

Adaptation de dose:

Situation:

Adjuvant  
 Néoadjuvant  
 Métastatique

Ligne de traitement:

+ ++ +++ - ? N 123 G D >>

**OBSERVATION MEDICALE - Compte-rendu de CHIMIO-ONCODIG**

**Cure de Chimiothérapie** | **INTERCURE (Toxicités observées après la précédente cure)** | **EXAMEN CLINIQUE**

**INTERCURE (Toxicités observées après la précédente cure)**

Aucune toxicité clinique

Vomissements:

Durée vomissements (jours):

Nausées:

Durée nausées (jours):

Stomatite:

Diarrhée (Camplo++):

**Neurotox : Si >= 14j : G2 I - Si >= 6 semaines fonctionnelle : G3 II**

Neurotox (Oxalplatine++):

Nombre de jours depuis précédente cure:

Si > 14 jours, motif:

Cutanée:

Alopécie:

Asthénie:

Neutropénie (SFU bolus++):

Thrombopénie:

Anémie:

Autre toxicité:

Durée (jours):

+ ++ +++ - ? N 123 G D >>

**OBSERVATION MEDICALE - Compte-rendu de CHIMIO-ONCODOG**

Cure de Chimiothérapie | INTERCURE (Toxicités observées après la précédente cure) | **EXAMEN CLINIQUE**

---

**EXAMEN CLINIQUE**

IP/DMS

Poids (k.g)  123

Taille (m)  123

Apyrétique  Oui  Non

PAS  123

PAD  123

Symptômes principaux: Examen clinique

Tolérance immédiate de la cure

**Sortie autorisée**

Oui  Non

Double du compte rendu remis au patient

Oui  Non

---

**EXAMENS BIOLOGIQUES (Bilan externe)**

Biologie  normale  anormale

PNN  123

Plaquettes  123

Hémoglobine  123

Bilirubine totale  123

Créatinine  123

Prochaine évaluation après cure n°  123

Commentaires





— Annexe B —

Questionnaire portant sur l'élicitation des poids de  
sévérité

# Scores de sévérité pour toxicités en oncologie.

L'objectif de ce questionnaire est de déterminer comment les oncologues prennent en compte les caractéristiques du patient (âge, sexe, effets indésirables, etc.) pour adapter les doses d'Irinotécan dans le traitement des cancers colorectaux, en pratique clinique.

Nous souhaitons recueillir votre opinion personnelle. Si vous souhaitez être informé de la suite du projet ou collaborer, vous pouvez nous recontacter.

Nous vous remercions de votre participation !

[sandrine.boulet-ext@aphp.fr](mailto:sandrine.boulet-ext@aphp.fr)

[annesophie.jannot@aphp.fr](mailto:annesophie.jannot@aphp.fr)

[sarah.zohar@inserm.fr](mailto:sarah.zohar@inserm.fr)

NEXT

Never submit passwords through Google Forms.

# Scores de sévérité pour toxicités en oncologie.

\* Required

## Questions

Pour chaque grade et chaque type de toxicités présentes dans le questionnaire DxCare (service HEPATO GASTRO, compte-rendu de CHIMIO-ONCODIG), veuillez attribuer un score entre 0 et 100. Ce score doit refléter l'importance que vous accordez à l'événement pour adapter votre stratégie thérapeutique (0 = nulle, 100 = considérable).

De la même manière, il vous est demandé d'attribuer un score entre 0 et 100 à un âge supérieur à 80 ans, à une perte de poids supérieure à 5 %, à différentes valeurs de bilirubine (> 35 micromol/L, > 50 micromol/L), aux grades du score OMS, aux différentes lignes de traitement et aux stades du cancer selon la classification TNM.

La dernière question concerne l'éventuel oubli d'une variable utilisée dans la réduction de dose.













Voyez-vous une autre variable que vous considérez dans l'ajustement de dose ?

Your answer

---

BACK

NEXT

Never submit passwords through Google Forms.

**Scores de sévérité pour toxicités en oncologie.**

**Merci pour votre aide !**

BACK

SUBMIT

Never submit passwords through Google Forms.

# Bibliographie

- [1] Yang Ai-Jun and Song Xin-Yuan. Bayesian variable selection for disease classification using gene expression data. *Bioinformatics*, 26(2) :215–222, January 2010.
- [2] Hirotugu Akaike. Information Theory and an Extension of the Maximum Likelihood Principle. In Emanuel Parzen, Kunio Tanabe, and Genshiro Kitagawa, editors, *Selected Papers of Hirotugu Akaike*, pages 199–213. Springer New York, New York, NY, 1998.
- [3] Australian Pancreatic Cancer Genome Initiative, Peter Bailey, David K. Chang, Katia Nones, Amber L. Johns, Ann-Marie Patch, Marie-Claude Gingras, David K. Miller, Angelika N. Christ, Tim J. C. Bruxner, Michael C. Quinn, Craig Nourse, L. Charles Murtaugh, Ivon Harliwong, Senel Idrisoglu, Suzanne Manning, Ehsan Nourbakhsh, Shivangi Wani, Lynn Fink, Oliver Holmes, Venessa Chin, Matthew J. Anderson, Stephen Kazakoff, Conrad Leonard, Felicity Newell, Nick Waddell, Scott Wood, Qinying Xu, Peter J. Wilson, Nicole Cloonan, Karin S. Kassahn, Darrin Taylor, Kelly Quek, Alan Robertson, Lorena Pantano, Laura Mincarelli, Luis N. Sanchez, Lisa Evers, Jianmin Wu, Mark Pinese, Mark J. Cowley, Marc D. Jones, Emily K. Colvin, Adnan M. Nagrial, Emily S. Humphrey, Lorraine A. Chantrill, Amanda Mawson, Jeremy Humphris, Angela Chou, Marina Pajic, Christopher J. Scarlett, Andreia V. Pinho, Marc Giry-Laterriere, Ilse Roman, Jaswinder S. Samra, James G. Kench, Jessica A. Lovell, Neil D. Merrett, Christopher W. Toon, Krishna Epari, Nam Q. Nguyen, Andrew Barbour, Nikolajs Zeps, Kim Moran-Jones, Nigel B. Jamieson, Janet S. Graham, Fraser Duthie, Karin Oien, Jane Hair, Robert Grützmann, Anirban Maitra, Christine A. Iacobuzio-Donahue, Christopher L. Wolfgang, Richard A. Morgan, Rita T. Lawlor, Vincenzo Corbo, Claudio Bassi, Borislav Rusev, Paola Capelli, Roberto Salvia, Giampaolo Tortora, Debabrata Mukhopadhyay, Gloria M. Petersen, Donna M. Munzy, William E. Fisher, Saadia A. Karim, James R. Eshleman, Ralph H. Hruban, Christian Pilarsky, Jennifer P. Morton, Owen J. Sansom, Aldo Scarpa, Elizabeth A. Musgrove, Ulla-Maja Hagbo Bailey, Oliver Hofmann, Robert L. Sutherland, David A. Wheeler, Anthony J. Gill, Richard A. Gibbs, John V. Pearson, Nicola Waddell, Andrew V. Biankin, and Sean M. Grimmond. Genomic analyses identify molecular subtypes of pancreatic cancer. *Nature*, 531(7592) :47–52, March 2016.
- [4] K. Bae and B. K. Mallick. Gene selection using a two-level hierarchical Bayesian model. *Bioinformatics*, 20(18) :3423–3430, December 2004.
- [5] R. D. Ball. Bayesian methods for quantitative trait loci mapping based on model selection : approximate analysis using the Bayesian information criterion. *Genetics*, 159(3) :1351–1364, November 2001.
- [6] B. Nebiyu Bekele and Peter F. Thall. Dose-Finding Based on Multiple Toxicities in a Soft Tissue Sarcoma Trial. *Journal of the American Statistical Association*, 99(465) :26–35, 2004.
- [7] Eric I. Benchimol, Liam Smeeth, Astrid Guttmann, Katie Harron, David Moher, Irene Petersen, Henrik T. Sorensen, Erik von Elm, Sinéad M. Langan, and RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLOS Medicine*, 12(10) :e1001885, October 2015.

- 
- [8] Donald A. Berry. Bayesian Methods in Phase III Trials. *Drug Information Journal*, 25(3) :345–368, July 1991.
- [9] Donald A. Berry and Dalene K. Stangl, editors. *Bayesian biostatistics*. Number v. 151 in Statistics, textbooks and monographs. M. Dekker, New York, 1996.
- [10] Scott M. Berry, editor. *Bayesian adaptive methods for clinical trials*. Number 38 in Chapman & Hall/CRC biostatistics series. CRC Press, Boca Raton, 2011. OCLC : ocn639940808.
- [11] Bhavana Bhatnagar, Steven Gilmore, Olga Goloubeva, Colleen Pelsler, Michelle Medeiros, Saranya Chumsri, Katherine Tkaczuk, Martin Edelman, and Ting Bao. Chemotherapy dose reduction due to chemotherapy induced peripheral neuropathy in breast cancer patients receiving chemotherapy in the neoadjuvant or adjuvant settings : a single-center experience. *SpringerPlus*, 3(1) :366, 2014.
- [12] John Blangero, Harald H. H. Goring, Jack W. Kent, Jeff T. Williams, Charles P. Peterson, Laura Almasy, and Thomas D. Dyer. Quantitative trait nucleotide analysis using Bayesian model selection. *Human Biology*, 77(5) :541–559, October 2005.
- [13] Howard D. Bondell and Brian J. Reich. Simultaneous regression shrinkage, variable selection and clustering of predictors with OSCAR. *Biometrics*, 64(1) :115–123, March 2008.
- [14] J. M. Brophy and L. Joseph. Placing trials in context using Bayesian analysis. GUSTO revisited by Reverend Bayes. *JAMA*, 273(11) :871–875, March 1995.
- [15] P. J. Brown, M. Vannucci, and T. Fearn. Multivariate Bayesian variable selection and prediction. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 60(3) :627–641, August 1998.
- [16] Aylin Caliskan, Joanna J. Bryson, and Arvind Narayanan. Semantics derived automatically from language corpora contain human-like biases. *Science*, 356(6334) :183–186, April 2017.
- [17] Bradley P. Carlin and Siddhartha Chib. Bayesian model choice via Markov Chain Monte carlo methods. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 57(3) :473–484, 1995.
- [18] C. M. Carvalho, N. G. Polson, and J. G. Scott. The horseshoe estimator for sparse signals. *Biometrika*, 97(2) :465–480, June 2010.
- [19] M.-H. Chen, J. G. Ibrahim, and C. Yiannoutsos. Prior elicitation, variable selection and Bayesian computation for logistic regression models. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 61(1) :223–242, February 1999.
- [20] Ming-Hui Chen, Joseph G. Ibrahim, Peter Lam, Alan Yu, and Yuanye Zhang. Bayesian design of noninferiority trials for medical devices using historical data. *Biometrics*, 67(3) :1163–1170, September 2011.
- [21] Ming-Hui Chen, Joseph G. Ibrahim, and Qi-Man Shao. Power prior distributions for generalized linear models. *Journal of Statistical Planning and Inference*, 84(1-2) :121–137, March 2000.
- [22] Ming-Hui Chen, Joseph G. Ibrahim, and Debajyoti Sinha. A New Bayesian Model for Survival Data with a Surviving Fraction. *Journal of the American Statistical Association*, 94(447) :909, September 1999.
- [23] Zhengjia Chen, Mark D. Krailo, Stanley P. Azen, and Mourad Tighiouart. A novel toxicity scoring system treating toxicity response as a quasi-continuous variable in Phase I clinical trials. *Contemporary Clinical Trials*, 31(5) :473–482, September 2010.
- [24] Hugh Chipman, Edward I. George, and Robert E. McCulloch. The Practical Implementation of Bayesian Model Selection. In *Institute of Mathematical Statistics Lecture Notes - Monograph Series*, pages 65–116. Institute of Mathematical Statistics, Beachwood, OH, 2001.
-

- [25] Peter Craven and Grace Wahba. Smoothing noisy data with spline functions : Estimating the correct degree of smoothing by the method of generalized cross-validation. *Numerische Mathematik*, 31(4) :377–403, December 1978.
- [26] Fulvio De Santis. Power Priors and Their Use in Clinical Trials. *The American Statistician*, 60(2) :122–129, May 2006.
- [27] Thérèse Deborde, Laurence Amar, Guillaume Bobrie, Nicolas Postel-Vinay, Christiane Battaglia, Amokrane Tache, Antoine Chedid, Mohamed-Malek Dhib, Gilles Chatellier, Pierre-François Plouin, Anita Burgun, Michel Azizi, and Anne-Sophie Jannot. Sex differences in antihypertensive treatment in France among 17856 patients in a tertiary hypertension unit. *Journal of Hypertension*, 36(4) :939–946, 2018.
- [28] Petros Dellaportas, Jonathan J. Forster, and Ioannis Ntzoufras. On Bayesian model and variable selection using MCMC. Technical report. *Statistics and Computing*, 12(1) :27–36, January 2002.
- [29] Jhonathan Pedroso Rigal dos Santos, Luiz Paulo Miranda Pires, Renato Coelho de Castro Vasconcellos, Gabriela Santos Pereira, Renzo Garcia Von Pinho, and Marcio Balestre. Genomic selection to resistance to *Stenocarpella maydis* in maize lines using DArTseq markers. *BMC Genetics*, 17(1), December 2016.
- [30] David M. Eddy, Victor Hasselblad, and Ross D. Shachter. *Meta-analysis by the confidence profile method : the statistical synthesis of evidence*. Statistical modeling and decision science. Academic Press, Boston, 1992.
- [31] Bradley Efron, Trevor Hastie, Iain Johnstone, and Robert Tibshirani. Least Angle Regression. *The Annals of Statistics*, 32(2) :407–451, 2004.
- [32] Monia Ezzalfani, Sarah Zohar, Rui Qin, Sumithra J. Mandrekar, and Marie-Cécile Le Deley. Dose-finding designs using a novel quasi-continuous endpoint for multiple toxicities. *Statistics in Medicine*, 32(16) :2728–2746, July 2013.
- [33] Jianqing Fan and Runze Li. Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties. *Journal of the American Statistical Association*, 96(456) :1348–1360, 2001.
- [34] Alessio Farcomeni. Bayesian Constrained Variable Selection. *Statistical Sinica*, 20 :1043–1062, 2010.
- [35] Jennifer Frankovich, Christopher A. Longhurst, and Scott M. Sutherland. Evidence-Based Medicine in the EMR Era. *New England Journal of Medicine*, 365(19) :1758–1759, November 2011.
- [36] Brooke L. Fridley. Bayesian variable and model selection methods for genetic association studies. *Genetic Epidemiology*, 33(1) :27–37, January 2009.
- [37] D. G. Fryback, J. O. Chinnis, and J. W. Ulvila. Bayesian cost-effectiveness analysis. An example using the GUSTO trial. *International Journal of Technology Assessment in Health Care*, 17(1) :83–97, 2001.
- [38] Edward I. George and Robert E. McCulloch. Variable Selection Via Gibbs Sampling. *Journal of the American Statistical Association*, 88(423) :881–889, 1993.
- [39] Edward I. George and Robert E. McCulloch. Approaches for Bayesian Variable Selection. *Statistica Sinica*, 7 :339–373, 1997.
- [40] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 2672–2680. Curran Associates, Inc., 2014.
- [41] Alessandra Gorini and Gabriella Pravettoni. An overview on cognitive aspects implicated in medical decisions. *European Journal of Internal Medicine*, 22(6) :547–553, December 2011.

- [42] Peter J. Green. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82(4) :711–732, 1995.
- [43] J. B. Greenhouse and L. Wasserman. Robust Bayesian methods for monitoring clinical trials. *Statistics in Medicine*, 14(12) :1379–1391, June 1995.
- [44] Shanti S. Gupta and James O. Berger, editors. *Statistical Decision Theory and Related Topics V*. Springer New York, New York, NY, 1994.
- [45] Tarek Haddad, Adam Himes, Laura Thompson, Telba Irony, Rajesh Nair, and on Behalf of MDIC Computer Modeling and Simulation Working Group Participants. Incorporation of stochastic engineering models as prior information in Bayesian medical device trials. *Journal of Biopharmaceutical Statistics*, 27(6) :1089–1103, November 2017.
- [46] Brian P. Hobbs, Bradley P. Carlin, Sumithra J. Mandrekar, and Daniel J. Sargent. Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials. *Biometrics*, 67(3) :1047–1056, September 2011.
- [47] James P. Hobert and George Casella. The Effect of Improper Priors on Gibbs Sampling in Hierarchical Linear Mixed Models. *Journal of the American Statistical Association*, 91(436) :1461, December 1996.
- [48] Nadine Houede, Peter F. Thall, Hoang Nguyen, Xavier Paoletti, and Andrew Kramar. Utility-based optimization of combination therapy using ordinal toxicity and efficacy in phase I/II trials. *Biometrics*, 66(2) :532–540, June 2010.
- [49] Joseph G. Ibrahim and Ming-Hui Chen. Prior Distributions and Bayesian Computation for Proportional Hazards Models. *Sankhya : The Indian Journal of Statistics, Series B (1960-2002)*, 60 :48–64, 1998.
- [50] Joseph G. Ibrahim and Ming-Hui Chen. Power Prior Distributions for Regression Models. *Statistical Science*, 15(1) :46–60, 2000.
- [51] Joseph G. Ibrahim, Ming-Hui Chen, Mani Lakshminarayanan, Guanghan F. Liu, and Joseph F. Heyse. Bayesian probability of success for clinical trials using historical data : J. G. IBRAHIM ET AL . *Statistics in Medicine*, 34(2) :249–264, January 2015.
- [52] Joseph G. Ibrahim, Ming-Hui Chen, and Steven N. MacEachern. Bayesian variable selection for proportional hazards models. *Canadian Journal of Statistics*, 27(4) :701–717, December 1999.
- [53] Joseph G Ibrahim, Ming-Hui Chen, and Debajyoti Sinha. On Optimality Properties of the Power Prior. *Journal of the American Statistical Association*, 98(461) :204–213, March 2003.
- [54] Joseph G. Ibrahim, Louise M. Ryan, and Ming-Hui Chen. Using Historical Controls to Adjust for Covariates in Trend Tests for Binary Data. *Journal of the American Statistical Association*, 93(444) :1282, December 1998.
- [55] Hemant Ishwaran and J. Sunil Rao. Spike and Slab Variable Selection : Frequentist and Bayesian Strategies. *The Annals of Statistics*, 33(2) :730–773, 2005.
- [56] Anne-Sophie Jannot, Eric Zapletal, Paul Avillach, Marie-France Mamzer, Anita Burgun, and Patrice Degoulet. The Georges Pompidou University Hospital Clinical Data Warehouse : A 8-years follow-up experience. *International Journal of Medical Informatics*, 102 :21–28, June 2017.
- [57] Frank G.A. Jansman, Dirk T. Sleijfer, Jules L.L.M. Coenen, Jacques C. De Graaf, and Jacobus R.B.J. Brouwers. Risk Factors Determining Chemotherapeutic Toxicity in Patients with Advanced Colorectal Cancer :. *Drug Safety*, 23(4) :255–278, 2000.
- [58] KE Lee, N Sha, ER Dougherty, M. Vannucci, and B. K. Mallick. Gene selection : a Bayesian variable selection approach. *Bioinformatics*, 19 :90–97, 2003.
-

- [59] Ki-Woong Kim, Young-Ju Suh, Woong-Yang Park, Jin-Hyeong Jhoo, Dong-Young Lee, Jong-Chul Youn, Kwang-Hyuck Lee, Jeong-Sun Seo, and Jong-Inn Woo. Choline acetyltransferase G +4 A polymorphism confers a risk for Alzheimer's disease in concert with Apolipoprotein E  $\epsilon$ 4. *Neuroscience Letters*, 366(2) :182–186, August 2004.
- [60] Christina M. R. Kitchen, Robert E. Weiss, Gang Liu, and Terri Wrin. HIV-1 viral fitness estimation using exchangeable on subsets priors and prior model selection. *Statistics in Medicine*, 26(5) :975–990, February 2007.
- [61] Lynn Kuo and Bani Mallick. Variable Selection for Regression Models. *Sankhya : The Indian Journal of Statistics, Series B (1960-2002)*, 60(1) :65–81, 1998.
- [62] Soonil Kwon, Dai Wang, and Xiuqing Guo. Application of an iterative Bayesian variable selection method in a genome-wide association study of rheumatoid arthritis. *BMC proceedings*, 1 Suppl 1 :S109, 2007.
- [63] Minjung Kyung, Jeff Gill, Malay Ghosh, and George Casella. Penalized regression, standard errors and Bayesian Lassos. *Bayesian Analysis*, 1, 2004.
- [64] Y. Li, C. Campbell, and M. Tipping. Bayesian automatic relevance determination algorithms for classifying gene expression data. *Bioinformatics*, 18(10) :1332–1339, October 2002.
- [65] David Madigan and Adrian E. Raftery. Model selection and accounting for model uncertainty in graphical models using Occam's window. *J Am Stat Assoc*, 89 :1535–1546, 1994.
- [66] Gilles Manceau, Sandrine Imbeaud, Raphaële Thiébaud, François Liébaert, Karine Fontaine, Francis Rousseau, Bérengère Génin, Delphine Le Corre, Audrey Didelot, Marc Vincent, Jean-Baptiste Bachet, Benoist Chibaudel, Olivier Bouché, Bruno Landi, Frédéric Bibeau, Karen Leroy, Frédérique Penault-Llorca, Jean-Luc Van Laethem, Pieter Demetter, Sabine Tejpar, Simona Rossi, Neda Mosakhani, Pia Österlund, Raija Ristamäki, Virinder Sarhadi, Sakari Knuutila, Valérie Boige, Thierry André, and Pierre Laurent-Puig. Hsa-miR-31-3p Expression Is Linked to Progression-free Survival in Patients with KRAS Wild-type Metastatic Colorectal Cancer Treated with Anti-EGFR Therapy. *Clinical Cancer Research*, 20(12) :3338–3347, June 2014.
- [67] Theo H. E. Meuwissen and Mike E. Goddard. Mapping multiple QTL using linkage disequilibrium and linkage analysis information and multitrait data. *Genetics, selection, evolution : GSE*, 36(3) :261–279, June 2004.
- [68] T. J. Mitchell and J. J. Beauchamp. Bayesian Variable Selection in Linear Regression. *Journal of the American Statistical Association*, 83(404) :1023–1032, 1988.
- [69] Ioannis Ntzoufras, Jonathan J. Forster, and Petros Dellaportas. Stochastic search variable selection for log-linear models. *Journal of Statistical Computation and Simulation*, 68(1) :23–37, December 2000.
- [70] Cheongeun Oh, Kenny Q. Ye, Qimei He, and Nancy R. Mendell. Locating disease genes using Bayesian variable selection with the Haseman-Elston method. *BMC genetics*, 4 Suppl 1 :S69, December 2003.
- [71] R. B. O'Hara and M. J. Sillanpää. A review of Bayesian variable selection methods : what, how and which. *Bayesian Analysis*, 4(1) :85–117, March 2009.
- [72] Chitu Okoli and Suzanne D. Pawlowski. The Delphi method as a research tool : an example, design considerations and applications. *Information & Management*, 42(1) :15–29, December 2004.
- [73] Trevor Park and George Casella. The Bayesian Lasso. *Journal of the American Statistical Association*, 103(482) :681–686, 2008.
- [74] Judea Pearl. *Causality : Models, Reasoning, and Inference*. Cambridge University Press, Cambridge, 2 edition, 2009.

- [75] Maya Petersen, Joshua Schwab, Susan Gruber, Nello Blaser, Michael Schomaker, and Mark van der Laan. Targeted Maximum Likelihood Estimation for Dynamic and Static Longitudinal Marginal Structural Working Models. *Journal of Causal Inference*, 2(2) :147–185, June 2014.
- [76] Adrian E. Raftery. Approximate Bayes factors and accounting for model uncertainty in generalized linear models. *Biometrika*, 83 :251–266, 1996.
- [77] Adrian E. Raftery, David Madigan, and Jennifer A. Hoeting. Bayesian Model Averaging for Linear Regression Models. *Journal of the American Statistical Association*, 92(437) :179–191, 1997.
- [78] Adrian E. Raftery, David Madigan, and Jennifer A. Hoeting. Model selection and accounting for model uncertainty in linear regression models. *J Am Stat Assoc*, 92 :179–191, 1997.
- [79] Charlotte Rietbergen, Irene Klugkist, Kristel J.M. Janssen, Karel G.M. Moons, and Herbert J.A. Hoijsink. Incorporation of historical data in the analysis of randomized therapeutic trials. *Contemporary Clinical Trials*, 32(6) :848–855, November 2011.
- [80] Veronika Rocková and Edward I. George. The Spike-and-Slab LASSO. *Journal of the American Statistical Association*, 113(521) :431–444, January 2018.
- [81] Parviz Safari, Syedeh Fatemeh Danyali, and Mehdi Rahimi. Bayesian inference for the genetic control of water deficit tolerance in spring wheat by stochastic search variable selection. *Environmental Science and Pollution Research*, 25(23) :23135–23142, August 2018.
- [82] Michael Schomaker, M.A. Luque-Fernandez, V. Leroy, and M.A. Davies. Using Longitudinal Targeted Maximum Likelihood Estimation in Complex Settings with Dynamic Interventions. 2018.
- [83] Gideon Schwarz. Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2) :461–464, March 1978.
- [84] Naijun Sha, Marina Vannucci, Mahlet G. Tadesse, Philip J. Brown, Ilaria Dragoni, Nick Davies, Tracy C. Roberts, Andrea Contestabile, Mike Salmon, Chris Buckley, and Francesco Falciani. Bayesian variable selection in multinomial probit models to identify molecular signatures of disease stage. *Biometrics*, 60(3) :812–819, September 2004.
- [85] Axel Skarman, Mohammad Shariati, Luc Jans, Li Jiang, and Peter Sorensen. A Bayesian variable selection procedure to rank overlapping gene sets. *BMC Bioinformatics*, 13(1) :73, 2012.
- [86] W S A Smellie. Methodology for constructing guidance. *Journal of Clinical Pathology*, 58(3) :249–253, March 2005.
- [87] Oleg Sofrygin, Zheng Zhu, Julie A. Schmittdiel, Alyce S. Adams, Richard W. Grant, Mark J. Laan, and Romain Neugebauer. Targeted learning with daily EHR data. *Statistics in Medicine*, April 2019.
- [88] David J. Spiegelhalter, Keith R. Abrams, and Jonathan P. Myles. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation : Spiegelhalter/Clinical Trials and Health-Care Evaluation*. John Wiley & Sons, Ltd, Chichester, UK, December 2003.
- [89] David J. Spiegelhalter, Laurence S. Freedman, and Mahesh K. B. Parmar. Bayesian Approaches to Randomized Trials. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 157(3) :357, 1994.
- [90] Sudeep Srivastava and Liang Chen. Comparison between the stochastic search variable selection and the least absolute shrinkage and selection operator for genome-wide association studies of rheumatoid arthritis. *BMC Proceedings*, 3(S7), December 2009.
- [91] Gertraud Stocker, Ulrich T. Hacker, Frédéric Fiteni, Jestinah John Mahachie, Arnaud D. Roth, Eric Van Cutsem, Marc Peeters, Florian Lordick, and Murielle Mauer. Clinical consequences
-

- of chemotherapy dose reduction in obese patients with stage III colon cancer : A retrospective analysis from the PETACC 3 study. *European Journal of Cancer*, 99 :49–57, August 2018.
- [92] Y. J. Suh, S. J. Finch, and N. R. Mendell. Application of a Bayesian method for optimal subset regression to linkage analysis of Q1 and Q2. *Genetic Epidemiology*, 21 Suppl 1 :S706–711, 2001.
- [93] Y. J. Suh, Ki-Woong Kim, Jin-Hyeong Jhoo, Dong-Young Lee, Jong-Chul Youn, YS Paek, IH Choo, JH Lee, and Jong-Inn Woo. Application of stochastic search variable selection to genetic association studies in common complex diseases. *Korean J Genet*, 28 :287–294, 2006.
- [94] Young Ju Suh, Kenny Q. Ye, and Nancy R. Mendell. A method for evaluating the results of Bayesian model selection : application to linkage analyses of attributes determined by two or more genes. *Human Heredity*, 55(2-3) :147–152, 2003.
- [95] Michael D. Swartz, Marek Kimmel, Peter Mueller, and Christopher I. Amos. Stochastic search gene suggestion : a Bayesian hierarchical model for gene mapping. *Biometrics*, 62(2) :495–503, June 2006.
- [96] Michael D. Swartz and Sanjay Shete. The null distribution of stochastic search gene suggestion : a Bayesian approach to gene mapping. *BMC proceedings*, 1 Suppl 1 :S113, 2007.
- [97] Mahlet G. Tadesse, Marina Vannucci, and Pietro Liò. Identification of DNA regulatory motifs using Bayesian variable selection. *Bioinformatics (Oxford, England)*, 20(16) :2553–2561, November 2004.
- [98] Zaixiang Tang, Yueping Shen, Xinyan Zhang, and Nengjun Yi. The Spike-and-Slab Lasso Generalized Linear Models for Prediction and Associated Genes Detection. *Genetics*, 205(1) :77–88, January 2017.
- [99] Jessica D. Tenenbaum, Paul Avillach, Marge Benham-Hutchins, Matthew K. Breitenstein, Erin L. Crowgey, Mark A. Hoffman, Xia Jiang, Subha Madhavan, John E. Mattison, Radhakrishnan Nagarajan, Bisakha Ray, Dmitriy Shin, Shyam Visweswaran, Zhongming Zhao, and Robert R. Freimuth. An informatics research agenda to support precision medicine : seven key areas. *Journal of the American Medical Informatics Association*, 23(4) :791–795, July 2016.
- [100] Peter F. Thall and Hoang Q. Nguyen. Adaptive randomization to improve utility-based dose-finding with bivariate ordinal outcomes. *Journal of Biopharmaceutical Statistics*, 22(4) :785–801, 2012.
- [101] Peter F Thall, Moreno Ursino, Véronique Baudouin, Corinne Alberti, and Sarah Zohar. Bayesian treatment comparison using parametric mixture priors computed from elicited histograms. *Statistical Methods in Medical Research*, 28(2) :404–418, February 2019.
- [102] Robert Tibshirani. Regression shrinkage and selection via the lasso : a retrospective : Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 73(3) :273–282, June 2011.
- [103] Mark J. van der Laan, Antoine Chambaz, and Sam Lendle. Online Targeted Learning for Time Series. In *Targeted Learning in Data Science*, pages 317–346. Springer International Publishing, Cham, 2018.
- [104] Mark J. van der Laan and I. Malenica. Robust Estimation of Data-Dependent Causal Effects based on Observing a Single Time-Series. 2018.
- [105] Mark J. van der Laan and Sherri Rose. *Targeted Learning*. Springer Series in Statistics. Springer New York, New York, NY, 2011.
- [106] Mark J. van der Laan and Sherri Rose. *Targeted Learning in Data Science : Causal Inference for Complex Longitudinal Studies*. Springer Series in Statistics. Springer International Publishing, Cham, 2018.



- [107] Joost van Rosmalen, David Dejardin, Yvette van Norden, Bob Löwenberg, and Emmanuel Lefsauffre. Including historical data in the analysis of clinical trials : Is it worth the effort ? *Statistical Methods in Medical Research*, 27(10) :3167–3182, October 2018.
- [108] V Viallefont, Adrian E. Raftery, and S Richardson. Variable selection and Bayesian model averaging in case-control studies. *Statistics in Medicine*, 20 :3215–3230, 2001.
- [109] Laura Visa, Paula Jiménez-Fonseca, Elena Asensio Martínez, Raquel Hernández, Ana Custodio, Manuel Garrido, Antonio Viudez, Elvira Buxo, Ignacio Echavarria, Juana Maria Cano, Ismael Macias, Montserrat Mangas, Eva Martínez de Castro, Teresa García, Felipe Alvarez Manceñido, Montes Ana Fernández, Aitor Azkarate, Federico Longo, Asunción Díaz Serrano, Carlos López, Alicia Hurtado, Paula Cerdá, Raquel Serrano, Aitziber Gil-Negrete, Alfonso Martín Carnicero, Paola Pimentel, Avinash Ramchandani, and Alberto Carmona-Bayonas. Efficacy and safety of chemotherapy in older versus non-older patients with advanced gastric cancer : A real-world data, non-inferiority analysis. *Journal of Geriatric Oncology*, 9(3) :254–264, May 2018.
- [110] CT Volinsky, David Madigan, Adrian E. Raftery, and RA Kronmal. Bayesian model averaging in proportional hazards models. *Appl Stat*, 46 :433–448, 1997.
- [111] Jon Wakefield and James Bennett. The Bayesian Modeling of Covariates for Population Pharmacokinetic Models. *Journal of the American Statistical Association*, 91(435) :917, September 1996.
- [112] Nengjun Yi, Varghese George, and David B. Allison. Stochastic search variable selection for identifying multiple quantitative trait loci. *Genetics*, 164(3) :1129–1138, July 2003.
- [113] Nengjun Yi and Shizhong Xu. Bayesian LASSO for quantitative trait loci mapping. *Genetics*, 179(2) :1045–1055, June 2008.
- [114] Ming Yuan and Yi Lin. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 68(1) :49–67, February 2006.
- [115] Z. Yuan, R. Chappell, and H. Bailey. The continual reassessment method for multiple toxicity grades : a Bayesian quasi-likelihood approach. *Biometrics*, 63(1) :173–179, March 2007.
- [116] Eric Zapletal, Nicolas Rodon, Natalia Grabar, and Patrice Degoulet. Methodology of integration of a clinical data warehouse with a clinical information system : the HEGP case. *Studies in Health Technology and Informatics*, 160(Pt 1) :193–197, 2010.
- [117] Lin Zhang, Veerabhadran Baladandayuthapani, Bani K. Mallick, Ganiraju C. Manyam, Patricia A. Thompson, Melissa L. Bondy, and Kim-Anh Do. Bayesian hierarchical structured variable selection methods with application to molecular inversion probe studies in breast cancer. *Journal of the Royal Statistical Society. Series C, Applied statistics*, 63(4) :595–620, August 2014.
- [118] Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 67(2) :301–320, April 2005.
-

## Résumé court

**Contexte :** La construction d'outils d'aide à la décision à destination des cliniciens nécessite de sélectionner, au préalable, les variables pertinentes pour modéliser la décision clinique. Pour cela, deux sources d'information peuvent être utilisées : les données observées recueillies dans les dossiers informatisés des patients, et l'expertise des médecins. Peu de méthodes permettent de combiner ces deux types d'information. L'originalité de ma thèse est donc de proposer des stratégies pour prendre en compte ces deux sources de données. Notre application porte sur la modélisation de la décision médicale de modification des doses d'irinotécan en fonction des caractéristiques des patients et des effets indésirables dans le traitement du cancer colorectal métastatique.

**Méthodes :** Les méthodes développées reposent sur des méthodes de sélection de variables bayésiennes dans lesquelles l'expertise est intégrée sous forme de poids associées à chaque variable et élicités par les experts. Le premier modèle est une adaptation d'une méthode de sélection de variables bayésienne, la méthode de *Stochastic Search Variable Selection* (SSVS), pour laquelle l'information experte est introduite dans la distribution a priori de l'indicatrice de sélection de variable. Le modèle a été appelé *Weights-Based SSVS* (WBS). Le deuxième modèle est fondé sur la méthode du *power prior* (PP) et permet de combiner des données simulées à partir des paramètres experts et les observations. La sélection de variables est elle aussi assurée par un modèle SSVS. Nous avons réalisé des simulations afin d'analyser les performances du premier modèle par rapport à celles des modèles LASSO et SSVS. Pour le cas d'usage, les données de soin des patients ayant reçu un protocole incluant de l'irinotécan à une dose théorique de  $180 \text{ mg/m}^2$  sur un cycle d'une durée théorique de 14 jours ont été extraites des dossiers patients informatisés de l'Hôpital Européen Georges Pompidou à Paris. Pour prendre en compte l'expertise clinique, nous avons construit un questionnaire en collaboration avec les oncologues dans lequel ils quantifient l'importance de chaque caractéristique des patients et de chaque grade de chaque type de toxicités sur la réduction de dose par l'intermédiaire de poids élicités prenant leurs valeurs entre 0 et 100. Enfin, nous avons appliqué ces modèles à des données de patients traités par irinotécan pour un cancer colorectal.

**Résultats :** Le modèle WBS présente de meilleures performances que les modèles uniquement basés sur les données observés. Cependant, ces performances dépendent des poids élicités par les cliniciens ainsi que de leur somme, qui doit donc être minutieusement calibrée en fonction du nombre de variables que l'on souhaite sélectionner. Le modèle basé sur le PP donne lui des performances similaires au modèle SSVS n'incluant que des données patient réellement observées. Toutefois, il ne sélectionne pas les mêmes variables et permet d'écartier des variables négligeables du point de vue expert tout en révélant celles qui, importantes mais peu fluctuantes dans les données, pourraient être omises. Concernant le cas d'usage, les poids élicités par les experts sont très variables et nos approches sélectionnent des variables pertinentes sur le plan clinique.

**Conclusion :** Cette thèse introduit deux méthodes Bayésiennes de sélection de variables combinant information experte élicitée et données observées. La première méthode permet d'améliorer les perfor-

mances prédictives sur des petits échantillons. La deuxième méthode permet de gérer des échantillons de plus grande taille, et de sélectionner automatiquement à la fois les variables importantes pour les experts et les variables utilisées en vie réelle. Les différences dans le choix des poids élicités par les oncologues montrent une certaine variabilité des pratiques professionnelles. La modélisation de la prise de décision médicale constitue une première étape dans la modélisation de la relation complexe qui lie caractéristiques du patients, doses de chimiothérapie, toxicités et survie.

**Mots clés :** Méthodes bayésiennes de sélection de variables ; *Stochastic Search Variable Selection* ; distribution a priori ; élicitation de poids de pertinence clinique ; mesures répétées ; dossiers de santé informatisés.

---

# Abstract

**Context :** Build decision support tools for clinicians first requires to select the relevant variables to model the clinical decision. For this, two sources of information can be used : the observed data collected in electronic health records, and the expertise of physicians. Few methods can combine these two types of information. The originality of my thesis is to propose strategies to take into account these two sources of data. Our use case focuses on modeling the medical decision to modify irinotecan doses based on patient characteristics and adverse events in the treatment of metastatic colorectal cancer.

**Methods :** The methods developed are based on Bayesian variable selection methods in which the expertise is integrated in the form of weights associated with each variable and elicited by the experts. The first model is an adaptation of a Bayesian variable selection method, the Stochastic Search Variable Selection (SSVS) method, for which expert information is introduced into the prior distribution of the variable selection indicator. The model was called Weights-Based SSVS (WBS). The second model is based on the power prior method (PP) and allows combining simulated data from expert parameters and observations. Variable selection is also provided by an SSVS model. We performed simulations to analyze the performances of the first model compared to the LASSO and SSVS models' ones. For the use case, the care data of the patients who received a protocol including irinotécan at a theoretical dose of  $180 \text{ mg/m}^2$  over a cycle of a theoretical duration of 14 days were extracted from the electronic health record of the Georges Pompidou European Hospital in Paris. To take into account the clinical expertise, we built a questionnaire in collaboration with the oncologists in which they quantify the importance of each patient characteristic and each grade of each toxicity type on dose reduction through weights taking their values between 0 and 100.

**Results :** The WBS model performs better than models based only on observed data. However, these performances depend on the weights elicited by the clinicians as well as their sum, which must therefore be carefully calibrated according to the number of variables that one wishes to select. The model based on the PP gives similar performances to the SSVS model including only patient data actually observed. However, it does not select the same variables and allows to exclude negligible variables from the expert point of view while revealing those which, important but rare in the data, could be omitted. Regarding the use case, the weights elicited by the experts are very variable and our approaches select clinically relevant variables.

**Conclusion :** This thesis introduces two Bayesian methods of variable selection combining elicited expert information and observed data. The first method improves predictive performance on small samples. The second method allows us to manage larger samples, and automatically select both important variables for experts and variables used in real life. The differences in the choice of weights elicited by oncologists show a certain variability of professional practices. The modeling of medical decision-making is a first step in modeling the complex relationship that links patient characteristics, chemotherapy doses, toxicities, and survival.

**Keywords :** Bayesian variable selection methods ; Stochastic Search Variable Selection ; prior distribution ; clinical relevance weights elicitation ; repeated measures ; electronic health records.

---