

ACADÉMIE DE VERSAILLES
UNIVERSITÉ DE VERSAILLES
SAINT-QUENTIN-EN-YVELINES
U.F.R. SIMONE VEIL - SANTÉ

ANNÉE 2022

N°

THÈSE
POUR LE DIPLÔME
D'ÉTAT DE DOCTEUR EN MÉDECINE
D.E.S. SANTÉ PUBLIQUE

Par Guillaume MULIER

Né le 22/01/1993 à Seclin

TITRE : Comparaison de schémas d'essais de phase II pour l'évaluation des nouveaux traitements en cancérologie

JURY :

Président : Sylvie CHEVRET, PU-PH

Directrice : Lucie BIARD, MCU-PH

Membre : Thomas APARICIO, PU-PH

Membre : David HAJAGE, MCU-PH

ACADÉMIE DE VERSAILLES
UNIVERSITÉ DE VERSAILLES
SAINT-QUENTIN-EN-YVELINES
U.F.R. SIMONE VEIL - SANTÉ

ANNÉE 2022

N°

THÈSE
POUR LE DIPLÔME
D'ÉTAT DE DOCTEUR EN MÉDECINE
D.E.S. SANTÉ PUBLIQUE
Par Guillaume MULIER
Né le 22/01/1993 à Seclin

TITRE : Comparaison de schémas d'essais de phase II pour l'évaluation des nouveaux traitements en cancérologie

JURY :

Président : Sylvie CHEVRET, PU-PH

Directrice : Lucie BIARD, MCU-PH

Membre : Thomas APARICIO, PU-PH

Membre : David HAJAGE, MCU-PH

Résumé

Introduction : Les nouvelles thérapeutiques en cancérologie telles que les immunothérapies et les thérapies ciblées remettent en cause les schémas usuels d'essais cliniques aux différentes phases de développement du médicament, du fait de leurs mécanismes d'action différents des chimiothérapies cytotoxiques. Nous avons pris l'exemple de la planification d'un essai de phase II évaluant une combinaison de nouvelles thérapies pour le traitement de l'adénocarcinome oesogastrique avancé nécessitant la définition d'un schéma adaptatif : fenêtre d'observation prolongée et évaluation conjointe et séquentielle de l'efficacité et de la toxicité.

Méthodes : Nous avons défini 6 schémas adaptatifs candidats par combinaison des approches suivantes : schéma bayésien permettant ou non un recrutement continu de patients, *monitoring* bayésien de la probabilité de toxicité séparé de l'évaluation d'efficacité ou évaluation conjointe des 2 critères selon un modèle multinomial, schéma de Simon en 2 étapes, intégration de données de toxicité externes. Ils ont été comparés par étude de simulation selon différents scénarios d'efficacité et de toxicité.

Résultats : Toutes les stratégies de schéma contrôlent le risque d'erreur de type I. Les schémas avec modélisation conjointe multinomiale de l'efficacité et de la toxicité ont une puissance supérieure. Les méthodes pondérées sur la fraction de suivi disponible lors des analyses intermédiaires permettent l'inclusion en continu des patients et ainsi une durée d'essai réduite en moyenne. La corrélation entre efficacité et toxicité influence les performances des schémas avec une augmentation du risque d'erreur de type I lorsque la corrélation est plus faible que planifiée.

Discussion : L'approche bayésienne avec modélisation multinomiale de l'efficacité et de la toxicité, avec pondération sur la fraction de suivi disponible lors des analyses intermédiaires semblent la plus avantageuse sur le plan des propriétés statistiques et de la durée de l'essai. De plus elle permet l'incorporation d'information extérieure via les distributions *a priori* utilisées pour l'inférence.

Mot-clefs : essais cliniques, phase II, schéma adaptatif, bayésien, cancer, immunothérapie.

Remerciements

Je tiens à remercier tout particulièrement Lucie Biard qui a accepté de m'encadrer sur cette thèse. Merci à elle pour son encadrement depuis le début de mon internat, sa pédagogie, ses conseils, sa disponibilité et sa patience car je peux être un élève très dissipé. Elle m'a enseigné beaucoup de ce que je connais maintenant.

Je remercie également vivement le professeur Sylvie Chevret pour l'accueil au sein du SBIM en tant qu'interne puis stagiaire, d'avoir accepté la présidence du jury, d'avoir été à l'écoute quand j'avais des questions, ainsi que de m'avoir encadré pour mon master 2.

Je remercie le professeur Thomas Aparicio d'avoir accepté que je travaille sur l'essai METINGAST et d'avoir accepté de faire partie du jury.

Je remercie le docteur Davis Hajage d'avoir accepté de faire partie du jury.

Merci à l'équipe du SBIM qui m'a accompagné, et particulièrement les doctorants (Anaïs, Valentin et Luana), Brigitte, Anouk et Déborah.

Mes remerciements vont également à mes parents qui m'ont poussé à aller sur Paris pour mes études et m'ont toujours soutenu, le reste de ma famille et mes amis.

Enfin, un merci spécial pour Axelle qui m'a beaucoup appris et avec qui je partage mes divagations sur R. Elle a toujours été de bon conseil et a aidé à relire cette thèse.

Table des matières

Résumé	2
Remerciements	3
Abréviations	8
1 Introduction	10
1.1 Les essais de phase précoce en oncologie	10
1.2 L'essai METIMGAST	13
1.3 Objectif	15
2 Méthodes	15
2.1 Les schémas d'essais étudiés	15
2.1.1 Monitoring bayésien de la toxicité	17
2.1.2 Le schéma de Simon en 2 étapes	18
2.1.3 Le schéma BOP ₂	20
2.1.4 Le schéma TOP	23
2.1.5 Apport de l'information sur la toxicité issue de la cohorte c-MET ⁺	25
2.2 Corrélation entre efficacité et toxicité	26
3 Article	27
3.1 Corps de l'article	27
3.2 Fichiers additionnels	39
3.2.1 Additional file 1 — Calibration of parameters for each design	39
3.2.2 Additional file 2 — Stopping rules for each design	45
3.2.3 Additional file 3 — Additional results	51
4 Conclusion	58
Bibliographie	60

Table des figures

Figure 1 : Percentage of conclusion of efficacy and no toxicity in the 10 scenarios with a positive correlation between efficacy and toxicity.....	38
Figure 2 : Percentage of conclusion of efficacy and no toxicity relative to the correlation between efficacy and toxicity in the setting.....	38
Figure S1 : Type I error rate and power on toxicity assessment, relative to parameters π_{tox} and τ in the decision rule	40
Figure S2 : Percentage of conclusion of efficacy and no toxicity in the 10 scenarios with a positive correlation between efficacy and toxicity for the different values of π_{tox} and τ	42
Figure S3 : Mean number of patients per trial in the 10 scenarios with a positive correlation between efficacy and toxicity for the different values of π_{tox} and τ	43
Figure S4 : Heatmap of the thresholds and a priori error rate relative to the correlation between efficacy and toxicity under H_0 and H_1 . The TOP-design was used with 2 interim analyses here	44
Figure S5 : Heatmap of the thresholds and a priori error rate relative to the correlation between efficacy and toxicity under H_0 and H_1 . (TOP with close monitoring of toxicity)	45
Figure S6 : Average number of patients per trial in the 10 scenarios with a positive correlation between efficacy and toxicity	51

Figure S7 : Proportion of early stopping in the 10 scenarios with a positive correlation between efficacy and toxicity	52
Figure S8 : Mean duration of a trial in the 10 scenarios with a positive correlation between efficacy and toxicity	53
Figure S9 : Percentage of conclusion of efficacy and no toxicity in the 10 scenarios with 6 sets of correlation between efficacy and toxicity	54
Figure S10 : Mean number of patients in the 10 scenarios with 6 sets of correlation between efficacy and toxicity	55
Figure S11 : Proportion of early stopping in the 10 scenarios with 6 sets of correlation between efficacy and toxicity	56
Figure S12 : Mean duration of a trial in the 10 scenarios with 6 sets of correlation between efficacy and toxicity	57

Table des tableaux

Table 1 : Bornes d'arrêt pour toxicité par l'approche PP_{tox}	18
Table 2 : Règles de décision pour les schémas de Simon pour l'essai METIMGAST	20
Table 3 : Exemple de table d'inclusion pour un schéma BOP_2 appliqué à l'évalua- tion jointe efficacité et toxicité dans la cohorte 1 de l'essai METIMGAST	23
Table 1 (article) : Simulation scenarios	37
Table S1 : Stopping rules for futility of $TOP_{eff} + PP_{tox}$ design	45
Table S2 : Stopping rules for toxicity of $TOP_{eff} + PP_{tox}$ design	46
Table S3 : Stopping rules for futility of Simon + PP_{tox} design	46
Table S4 : Stopping rules for futility of $BOP_{eff} + PP_{tox}$ design	46
Table S5 : Stopping rules for futility of $TOP_{eff/tox}$ design with 2 analyses	47
Table S6 : Stopping rules for toxicity of $TOP_{eff/tox}$ design with 2 analyses	47
Table S7 : Stopping rules for futility of $TOP_{eff/tox}$ design with close monitoring of toxicity	48
Table S8 : Stopping rules for toxicity of $TOP_{eff/tox}$ design with close monitoring of toxicity (1/2)	49
Table S9 : Stopping rules for toxicity of $TOP_{eff/tox}$ design with close monitoring of toxicity (2/2)	50

Abréviations

AMM : Autorisation de Mise sur le Marché;

ANSM : Agence Nationale de Sécurité du Médicament et des produits de santé;

APHP : Assistance Publique - Hôpitaux de Paris;

BOIN : Bayesian Optimal INterval design;

BOP₂ : Bayesian Optimal design for Phase II clinical trials;

CNIL : Commission Nationale de l'Informatique et des Libertés;

CPP : Comité de Protection des Personnes;

CRM : Continual Reassessment Method;

Dir() : Loi de Dirichlet;

DMT : Dose Maximale Tolérée;

DRCI : Délégation à la Recherche Clinique et à l'Innovation;

EMA : European Medicines Agency;

ESS : Effective Sample Size;

FDA : Food and Drug Administration;

ITK : Inhibiteur de Tyrosine Kinase;

OBD : Optimal Biologic Dose;

MAMS : Multi-Arm Multi-stage;

c-MET : Récepteur du Hepatocyte Growth Factor;

Multinom() : Loi multinomiale;

mTPI : modified Toxicity Probability Interval design;

NMP : Nombre Moyen de Patients;

PAE : Probabilité d'Arrêt d'un Essai;

PAP : Probabilité D'arrêt Précoce;

PAT : Probabilité d'Arrêt Tardif;

PDL₁ : Programmed cell Death Ligand 1;

PET : Probability of Early Termination;

PIPE : Product of Independant beta Probabilities dose Escalation;

PLT : Probability of Late Termination;

RP₂D : Recommended Phase 2 Dose;

TITE-CRM : Time-to-Event CRM;

TOP : Time-to-event BOP₂.

I Introduction

I.1 Les essais de phase précoce en oncologie

Le développement d'un médicament, de l'identification d'une molécule d'intérêt jusqu'à sa commercialisation, dure approximativement 15 ans et est découpé en plusieurs phases successives (1; 2). Tout d'abord, la phase préclinique, avec les expérimentations *in vitro* et sur les modèles animaux, vise à détecter les molécules d'intérêt, à la fois sur des aspects d'efficacité que de toxicité. Ensuite survient la partie clinique, c'est-à-dire chez l'Homme, divisée traditionnellement en 4 phases : la phase I pour évaluer la toxicité du médicament et déterminer une(des) dose(s) tolérable(s) pour la suite de son développement, la phase II pour commencer à évaluer l'efficacité du médicament, parfois en comparant différentes doses ayant une toxicité tolérable, la phase III pour confirmer l'efficacité du médicament qui précède la demande d'autorisation de mise sur le marché (on parle alors d'essai d'enregistrement) et la phase IV, après l'autorisation de mise sur le marché (AMM), évaluant le médicament en conditions de vie réelle, avec la pharmacovigilance. En règle générale, le terme "phase précoce" concerne des essais de phase I et II.

De par la sévérité de la maladie et la toxicité des traitements, les essais de phase I en oncologie sont réalisés sur des patients, tandis qu'ils sont habituellement réalisés sur des volontaires sains dans les autres cas. Les schémas d'essais de phase I reposent sur l'inclusion de cohortes de patients (classiquement entre 1 et 3 patients par cohorte) à qui une dose est attribuée parmi un ensemble de doses candidates. Le but est d'identifier la dose maximale tolérée (DMT) en traitant le moins possible de patients à des doses toxiques, en escaladant (ou désescaladant) les doses selon les résultats des cohortes précédentes. Les schémas de phase I sont classés actuellement en 3 grands types :

- les schémas *rule-based* qui s'appuient sur des règles de décision sans spécifier *a priori* la forme de la relation dose-toxicité ; le schéma *rule-based* le plus utilisé est le schéma 3 + 3 (3),
- les schémas *model-based* qui s'appuient sur une modélisation paramétrique de la rela-

tion dose-toxicité pour construire les règles de décision pendant l'essai et déterminer la DMT. Le premier proposé, en 1990, est la *continual reassessment method* (CRM) (4); de nombreux autres ont été développés depuis,

- les schémas *model-assisted*, à mi-chemin entre les 2 précédents, qui ne comportent pas de paramétrisation de la relation dose toxicité mais utilisent tout de même, en règle générale, une modélisation paramétrique de la probabilité de toxicité pour chaque dose candidate séparément (exemple : schémas mTPI (*modified toxicity probability interval design*) (5), PIPE (*product of independent beta probabilities dose escalation*) (6), BOIN (*Bayesian Optimal Interval design*) (7)).

Les schémas de phase II ensuite ont pour but d'écarter les médicaments non efficaces. Au cours du développement d'un médicament, les premiers essais de phase II sont classiquement des essais simple bras sur un petit effectif de patients, avec le traitement administré à la dose déterminée en phase I (DMT ou RP₂D, *recommended phase II dose*), et dont l'activité est comparée à une valeur de référence d'efficacité, comme par exemple le schéma de Simon (8). Une particularité en cancérologie est la toxicité non négligeable des traitements; ainsi, à la place d'une évaluation de l'efficacité seule, une évaluation conjointe de l'efficacité et de la toxicité est parfois préférée (9; 10; 11).

Les essais de phase III en cancérologie demeurent pourtant encore trop souvent négatifs et l'approche d'une molécule à la fois manque de spécificité pour écarter les molécules non efficaces (12). Cette approche dite *single arm* peut en effet aboutir à un *treatment-trial confounding* (confusion entre l'effet réel du traitement et un effet "essai") (13) et pour remédier à cela, d'autres approches ont été proposées comme les essais randomisés contre un groupe contrôle dès la phase II (13) ou l'approche multi-bras dite MAMS (= Multi-Arm Multi-Stage), avec plusieurs traitements évalués (12). Cependant, le développement de thérapeutiques en oncologie s'écarte de plus en plus fréquemment de ce cadre général.

D'une part la gravité de la maladie, combinée à un besoin thérapeutique demeurant important, peut entraîner une accélération du développement des thérapies. On a ainsi vu apparaître des schémas d'essais de phase I/II (14; 15; 16; 17), II/III (18; 19), I-II/III (20), ainsi

que l'essor des schémas séquentiels ou adaptatifs avec des règles d'arrêt précoce participant à l'accélération du développement (21). Parallèlement, les autorités ont pris en compte cette évolution : l'*European Medicines Agency* (EMA) ainsi que la *Food and Drug Administration* (FDA) ont publié des recommandations sur l'utilisation des schémas adaptatifs pour les essais cliniques (22; 23; 24), et des procédures de développement de médicament accélérées, dites *Fast Track*, ont été mises en place pour réduire le temps de développement (25; 26; 27).

D'autre part, les caractéristiques pharmacologiques des nouveaux traitements en oncologie, tels que les immunothérapies et les thérapies ciblées, s'écartent de celles des chimiothérapies cytotoxiques sur lesquelles repose historiquement la structure de l'évaluation précoce des médicaments (28). En premier lieu, les relations dose-activité ne sont plus nécessairement strictement monotones croissantes (16; 29). La dose maximale tolérée n'est alors pas nécessairement optimale, en termes de balance toxicité-efficacité. La distinction entre phases I et II pour l'évaluation d'abord de la toxicité puis de l'efficacité est alors souvent remplacée par l'évaluation conjointe de la toxicité et de l'efficacité (10; 17). Au lieu de prendre la DMT en phase 2 comme dose recommandée en phase II (RP_{2D}), on parle de Dose Optimale Biologique (OBD) définie comme le dose non toxique avec la plus grande efficacité (30).

De plus, ces nouveaux traitements ont un schéma d'administration différent des chimiothérapies cytotoxiques. En règle générale, ils s'administrent sur un mode continu et prolongé, contrairement aux chimiothérapies cytotoxiques qui sont usuellement administrées par cycles, en nombre limité. Cela conduit à des réponses et des toxicités plus tardives, donc les fenêtres d'observations ont dû être adaptées en conséquence dans les essais cliniques et des critères de jugements de substitution observables plus précocement (par exemple la survie sans récurrence au lieu de la survie globale) sont utilisés (31).

La section suivante présente l'essai qui a motivé ce travail et illustre ces différentes problématiques méthodologiques et statistiques pour l'évaluation des médicaments en cancérologie en phase II.

1.2 L'essai METIMGAST

L'essai METIMGAST (investigateur coordonateur : Pr Thomas Aparicio, Hôpital Saint Louis, Paris, France ; promoteur : Assistance Publique - Hôpitaux de Paris) est un essai de phase II, non comparatif, évaluant l'efficacité et la toxicité de l'association spartalizumab (un anticorps monoclonal inhibiteur de point de contrôle immunitaire, ciblant le *Programmed cell Death Ligand 1* (PDL1)) et capmatinib (un Inhibiteur de Tyrosine Kinase (ITK) ciblant le récepteur du *Hepatocyte Growth Factor* (c-MET)) chez des patients atteints d'un adénocarcinome de la jonction oesogastrique métastatique ou localement avancé ayant reçu au moins une ligne de chimiothérapie à base de sel de platine et de fluoropyrimidine avec une progression documentée sous chimiothérapie (NCT05135845).

L'adénocarcinome de la jonction oesogastrique est généralement considéré comme faisant partie des cancers gastriques (32) qui représentent la 2^{ème} localisation parmi les cancers digestifs, et la 5^{ème} cause de cancer dans le monde. C'est aussi le 2^{ème} cancer digestif le plus meurtrier (33). Le pronostic de l'adénocarcinome de la jonction oesogastrique reste très sévère, même après le traitement de 1^{ère} ligne qui est la résection chirurgicale associée à une polychimiothérapie (si le cancer est localement avancé), avec une médiane de survie globale de 4 ans si la tumeur est localisée et moins de 1 an pour les tumeurs localement avancées (34; 35). Les traitements de 2^{nde} ligne n'améliorent que peu la survie des patients réfractaires (36; 37; 38).

Des nouveaux traitements par anti-PDL1 ont récemment été évalués dans les cancers gastriques et ont montré des résultats prometteurs (39; 40). De plus, certaines études *in vitro* et sur modèles animaux suggèrent une action synergique des anti-PDL1 lorsqu'associés avec des inhibiteurs du c-MET (41; 42). Le récepteur c-MET, lorsqu'il est exprimé de façon anormale par la tumeur à cause du proto-oncogène MET, représente une cible thérapeutique. De plus, une autre piste d'action est étudiée pour les inhibiteurs c-MET qui est indépendante du proto-oncogène MET de la tumeur. Un effet thérapeutique pourrait être médié par une action d'inhibition du recrutement local des polynucléaires neutrophiles et une augmentation du recrutement des lymphocytes T CD8+.

L'objectif de l'essai METIMGAST est d'évaluer l'efficacité et la toxicité de l'association spartalizumab et capmatinib pour le traitement de l'adénocarcinome oesogastrique avancé en 2nde ligne. Dans la mesure où il est estimé qu'environ 10% des cancers gastriques et des cancers de la jonction oesogastrique ont une amplification du proto-oncogène MET (43; 44) et que cet amplification pourrait être une source d'hétérogénéité pour l'efficacité des inhibiteurs de c-MET (45), le schéma de l'essai prévoit 2 cohortes : la cohorte principale qui est composée de patients ne présentant pas d'amplification c-MET (c-MET⁻) et une cohorte annexe, minoritaire, composée de patients porteurs de l'amplification c-MET (c-MET⁺). La cohorte principale c-MET⁻ servira pour l'analyse principale de l'efficacité; la cohorte minoritaire c-MET⁺, quant à elle, sera analysée à visée exploratoire pour l'efficacité. Par contre, il n'y a pas de différence attendue de toxicité donc l'analyse de la toxicité prend en compte les 2 cohortes.

Le schéma prévu est décrit en détail à la section 2.1. Brièvement, un schéma adaptatif de phase II avec une analyse intermédiaire d'efficacité, a été retenu afin d'évaluer l'activité anti-tumorale de la combinaison. Compte-tenu de la toxicité connue du capmatinib, un monitoring rapproché de la toxicité a été mis en place en parallèle de l'évaluation d'efficacité, avec des règles d'arrêt précoce. Un effectif total maximal de 90 patients est planifié : l'effectif estimé de la 1^{ère} cohorte est de 81 patients (cMET⁻), celui de la 2^{ème} cohorte de 9 patients (cMET⁺).

En tant que recherche interventionnelle impliquant la personne humaine, l'essai METIMGAST est porté par l'Assistance Publique Hôpitaux de Paris (AP-HP) en tant que promoteur, avec un financement institutionnel par le programme CLIPP Molécules Innovantes de l'Institut National du Cancer. Conformément à l'article L1123-6 du Code de la Santé Publique, l'essai a été soumis au Comité de Protection de Personnes (CPP) qui a rendu un avis favorable initialement le 14/06/2021, et à l'ANSM, qui a donné son autorisation initiale le 21/07/2021. Enfin, l'essai suit la méthodologie de référence MR-001 de la CNIL (Commission Nationale de l'Informatique et des Libertés). L'essai est en cours actuellement, en période d'inclusion, le premier patient ayant été inclus en mars 2022.

1.3 Objectif

L'objectif est de définir et étudier les propriétés statistiques de différents schémas d'essais de phase II simple-bras adaptés à l'évaluation des nouveaux traitements en cancérologie : fenêtre d'observation prolongée, analyses intermédiaires avec règles d'arrêt précoces, évaluation conjointe de l'efficacité et de la toxicité.

Nous nous sommes appuyés sur la planification de l'essai METIMGAST (NCT05135845), qui évalue une nouvelle combinaison thérapeutique pour le traitement de l'adénocarcinome de la jonction oesogastrique, localement avancé ou métastatique, chez l'adulte.

2 Méthodes

2.1 Les schémas d'essais étudiés

Nous avons comparé le schéma prévu pour l'essai METIMGAST à 5 autres schémas adaptés au projet METIMGAST à l'aide d'une étude de simulations (46) afin d'évaluer leurs propriétés.

Nous avons combiné différentes approches proposées dans la littérature pour l'évaluation de l'efficacité et de la toxicité de façon à obtenir des schémas d'essai permettant de répondre aux spécificités suivantes :

- évaluation de l'efficacité du traitement avec un schéma de phase II non contrôlé,
- règle d'arrêt précoce pour futilité, avec une analyse intermédiaire,
- règle d'arrêt précoce pour toxicité, avec surveillance (*monitoring*) rapprochée
- fenêtre d'observation de 6 mois pour l'évaluation de l'efficacité, pour une fréquence d'inclusion attendue de 5 patients par mois.

Les 6 schémas comparés sont présentés en détail dans les sections suivantes; de façon résumée, ils sont les suivants :

- le schéma Time to event Bayesian Optimal Phase II design for clinical trials (TOP) (47) + suivi bayésien de la toxicité à l'aide de la distribution *a posteriori* de la probabilité de toxicité (PP_{tox}) (11; 48) (il s'agit du schéma prévu dans l'essai METIMGAST) : un

- schéma TOP pour l'évaluation de l'efficacité avec 1 analyse intermédiaire à 30 patients et l'analyse finale à 81 patients et un monitoring de la toxicité par l'approche suivant la probabilité *a posteriori* de toxicité, PP_{tox} , à 5, 10, 15, 20, 30, 40, 50, 60, 70, 80 et 90 patients dans les 2 groupes de patients, c-MET+ et c-MET- ;
- le schéma Bayesian Optimal Phase II design for clinical trials (BOP₂) (49) + PP_{tox} : il s'agit d'une approche similaire à la précédente où le schéma TOP est remplacé par le schéma BOP₂ pour l'évaluation de l'efficacité, en gardant les mêmes effectifs pour les analyses. La différence avec le schéma précédent réside dans le fait que, pour chaque cohorte de patients inclus, il est nécessaire d'interrompre les inclusions et attendre que l'observation du critère de jugement soit complète pour tous les patients inclus jusqu'alors avant de faire l'analyse (ce qui peut conduire à devoir attendre la fin de la fenêtre d'observation du dernier inclus, c'est-à-dire 6 mois dans le cas de METIM-GAST), pour ensuite d'autoriser la reprise du recrutement ou d'arrêter l'essai ;
 - le schéma de Simon en 2 étapes (8) pour l'évaluation de l'efficacité à 30 et 81 patients dans la cohorte 1, associé à un monitoring rapproché de la toxicité par l'approche PP_{tox} identique à celui des 2 schémas précédents ;
 - le schéma TOP implémenté dans sa version avec critères multiples pour l'évaluation jointe de l'efficacité et la toxicité dans la cohorte I (c-MET-), avec 2 analyses, intermédiaire à 30 patients et finale à 81 ;
 - le schéma TOP implémenté dans sa version avec critères multiples, efficacité et toxicité, adapté pour des analyses intermédiaires de toxicité plus fréquentes (monitoring rapproché, avec 10 analyses à 5, 10, 15, 20, 30, 40, 50, 60, 70 et 81 patients) mais 2 analyses d'efficacité (à 30 et 81 patients) ;
 - le schéma TOP implémenté dans sa version avec critère multiples, efficacité et toxicité, avec des analyses intermédiaires plus fréquentes pour toxicité tout comme le schéma précédent, avec une prise en compte de l'information apportée par la cohorte 2 (c-MET+) pour l'évaluation de la toxicité.

2.1.1 Monitoring bayésien de la toxicité

Plusieurs auteurs ont proposé de surveiller le risque de toxicité dans les essais de phase II à l'aide d'une approche bayésienne avec la distribution *a posteriori* de la probabilité de toxicité (11; 50; 48). La notation PP_{tox} est utilisée pour cette méthode dans le cadre du présent travail.

Les toxicités parmi les patients de l'essai peuvent être modélisées par une distribution béta-binomiale. La surveillance (*monitoring*) s'appuie alors sur la distribution *a posteriori* de la probabilité de toxicité p_{tox} (loi béta) obtenu en combinant les connaissances *a priori* sur ce risque et les données observées de l'essai, D_n . Des règles d'arrêt peuvent être définies en considérant la probabilité *a posteriori* que p_{tox} dépasse un certain seuil π_{tox} , c'est-à-dire que le risque de toxicité soit inacceptable : $P(p_{\text{tox}} > \pi_{\text{tox}} | D_n)$. Si cette probabilité est trop importante (typiquement supérieure à 0,95 ou 0,98) alors l'essai peut être interrompu. Pour l'essai METIMGAST, la règle d'arrêt de toxicité s'écrit comme suit :

$$P(p_{\text{tox}} > 0,25 | D_n) > 0,95.$$

Les valeurs de 25% de risque de toxicité maximale tolérée et du seuil de 0,95 ont été définies à partir des connaissances des cliniciens et des propriétés statistiques du schéma ainsi obtenu. Par simulations, il a en effet été évalué que si la probabilité de toxicité est réellement limitée, par exemple inférieure à 0,20 ($p_{\text{tox}} \leq 0.20$), la probabilité d'un essai positif en combinant le TOP et le monitoring de la toxicité reste aux alentours de 5%. Inversement, lorsque $p_{\text{tox}} = 0,30$, la proportion d'arrêt précoce de l'essai est d'environ 49%, et de 95% lorsque $p_{\text{tox}} = 0,40$.

Pour le présent travail, nous avons modifié cette règle d'arrêt, telle que : $P(p_{\text{tox}} > 0,20 | D_n) > 0,95$. Le choix d'un seuil de 20% pour un risque de toxicité inacceptable au lieu de 25% permet d'avoir un taux de faux positifs (essai aboutissant, à tort, à une conclusion de traitement efficace et non toxique) moindre (évaluation par simulations). Les règles d'arrêt produites par cette règle de décision pour l'essai METIMGAST sont présentées dans le tableau 1.

Les décomptes de toxicités ont été considérés à la fin de la période d'observation de tous

les patients suivis pour le schéma BOP+PP_{tox} et au moment de l'analyse pour efficacité ou après le dernier patient inclus pour le schéma TOP+PP_{tox}.

TABLE 1 – Bornes d'arrêt pour toxicité par l'approche PP_{Tox} : $P(p_{\text{tox}} > 0,25|D_n) > 0,95$ (entre parenthèses les seuils pour la règle d'arrêt modifiée, $P(p_{\text{tox}} > 0,20|D_n) > 0,95$).

Analyse	Nombre de patients inclus	Arrêt pour toxicité si $\geq \dots$ toxicités
1	5	3 (3)
2	10	5 (5)
3	15	7 (6)
4	20	9 (7)
5	30	12 (10)
6	40	15 (13)
7	50	18 (15)
8	60	21 (17)
9	70	24 (20)
10	80	27 (22)
11 (finale)	90	30 (25)

2.1.2 Le schéma de Simon en 2 étapes

Simon (8) a proposé en 1989 un schéma de phase II simple bras en 2 étapes permettant d'arrêter précocement un essai en présence d'un signal de non efficacité (arrêt pour futilité).

Le schéma comporte une analyse intermédiaire et une analyse finale. Pour calculer les règles de décision à ces analyses, on formule 2 hypothèses : l'hypothèse nulle $H_0 : p < p_0$ et l'hypothèse alternative $H_1 : p \geq p_1$, avec p la probabilité d'efficacité du traitement, p_0 un niveau non intéressant d'efficacité, et p_1 un niveau prometteur et désirable d'efficacité. Les règles de décision à chacune des 2 étapes sont les suivantes : sachant n_1 et n les effectifs de patients inclus aux analyses intermédiaire et finale respectivement, s'il y a moins de r_1 réponses parmi n_1 patients à l'analyse intermédiaire ou moins de r réponses parmi n à l'analyse finale, le traitement est considéré comme non efficace et l'essai est arrêté précocement dans le cas de l'analyse intermédiaire.

Le schéma s'appuie sur les probabilités conditionnelles et la loi binomiale. A l'analyse intermédiaire, la probabilité de s'arrêter précocement (PAP) pour non efficacité peut s'exprimer de la façon suivante : $\text{PAP} = B(r_1|p, n_1)$ avec B la fonction de répartition de la loi binomiale,

r_1 le nombre maximum de réponses lors de la 1^{ère} analyse pour s'arrêter pour futilité, n_1 le nombre de patients recrutés lors de cette analyse et p la probabilité de réponse pour le traitement étudié. De façon analogue, lors de la 2^{ème} analyse, on peut calculer la probabilité de conclusion de futilité (PAT) sachant qu'on ne s'est pas arrêté à l'analyse intermédiaire :
$$\text{PAT} = \sum_{x=r_1+1}^{\min(n_1, r)} b(x|p, n_1)B(r-x|p, n_2)$$
 avec r le nombre maximum de réponses lors de l'analyse finale pour conclure à la non-efficacité, n le nombre de patients recrutés dans l'essai, $n_2 = n - n_1$ et b la fonction de densité de la loi binomiale. La probabilité d'arrêt d'un essai (PAE) s'obtient alors en sommant ces 2 probabilités.

Dans le cadre de ce schéma, l'erreur de type I α est donc définie comme la PAE lorsque $p = p_0$, et la puissance $1 - \beta$ comme la PAE lorsque $p = p_1$. On peut également définir le nombre moyen de patients (NMP) = $\text{PAP} \times n_1 + (1 - \text{PAP}) \times n = n_1 + (1 - \text{PAP}) \times n_2$.

Pour la planification d'un essai donné, le schéma de Simon s'obtient en spécifiant les hypothèses p_0 et p_1 sur l'efficacité du traitement et les risques α et β . Par recherche de seuil parmi les valeurs possibles de r_1, r, n_1 et n (51), 2 principaux schémas répondant aux contraintes sur α et β sont habituellement considérés de façon préférentielle :

- le schéma optimal qui est le schéma qui minimise le nombre moyen de patients lorsque l'efficacité est égale à p_0 ;
- le schéma minimax qui est le schéma qui minimise le nombre maximum de patients recrutés n lorsque l'efficacité est égale à p_0 .

Les règles de décision de ces 2 schémas, appliqués à l'essai METIMGAST pour l'analyse d'efficacité dans la cohorte principale c-MET –, sont présentées à titre indicatif dans le tableau 2. Dans le but de comparer les schémas à effectifs identiques pour l'essai METIMGAST, nous avons forcé les effectifs utilisés dans les autres schémas pour l'évaluation de l'efficacité aux 2 analyses du schéma de Simon ($n_1 = 30$ et $n = 81$) et déterminé le schéma minimisant le nombre moyen de patients sous H_0 avec ces effectifs fixés (dénomé "Custom" dans le tableau 2).

TABLE 2 – Règles de décision pour les schémas de Simon pour l'essai METIMGAST

Schéma	r_1	n_1	r	n	EN(p_0)	PET(p_0)
Optimal	5	30	17	82	45.05	0.71
Minimax	6	42	14	64	51.80	0.55
Custom	4	30	17	81	54.30	0.52

2.1.3 Le schéma BOP₂

Le schéma BOP₂ (49) est un schéma d'essai de phase II simple bras permettant l'ajout d'analyses intermédiaires. De plus, il permet de modéliser sous la même approche des critères binaires uniques (exemple : efficacité, toxicité) ou multiples (exemple : 2 co-critères d'efficacité, efficacité et toxicité). Ce schéma est par essence bayésien avec l'estimation de probabilités *a posteriori* conditionnellement aux données recueillies (52) à partir desquelles sont définies les règles d'arrêt.

Avec ce schéma, nous nous intéressons à m paramètres binaires différents résultant en $J = 2^m$ différentes possibilités de résultats. Par exemple, dans le contexte de ce travail, nous prenons $m = 2$ critères, à savoir efficacité et toxicité (X_E et X_T qui valent 1 si respectivement présence d'efficacité ou toxicité et 0 sinon), ce qui aboutit donc à $2^2 = 4$ résultats possibles. On dénotera ainsi $Y = (Y_1, Y_2, Y_3, Y_4)$ la variable aléatoire multinomiale qui décrit les résultats possibles :

$$Y \sim \text{Multinomial}(n, \theta_1, \theta_2, \theta_3, \theta_4).$$

Les 4 résultats possibles sont $Y_1 = \mathbb{1}(X_E = 1 \cap X_T = 1)$: "efficacité et toxicité", $Y_2 = \mathbb{1}(X_E = 1 \cap X_T = 0)$: "efficacité et pas de toxicité", $Y_3 = \mathbb{1}(X_E = 0 \cap X_T = 1)$: "pas d'efficacité et toxicité" et $Y_4 = \mathbb{1}(X_E = 0 \cap X_T = 0)$: "pas d'efficacité et pas de toxicité", avec $\mathbb{1}()$ la fonction indicatrice.

Pour pouvoir planifier le schéma, nous avons besoin de 2 hypothèses :

- une hypothèse d'inefficacité et de toxicité : $Y \sim \text{Multinomial}(n, \theta_1^0, \theta_2^0, \theta_3^0, \theta_4^0)$ avec θ_k^0 qui représentent les paramètres de la loi pour laquelle le traitement sera considéré non efficace et trop toxique ;
- une hypothèse d'efficacité et de non toxicité : $Y \sim \text{Multinomial}(n, \theta_1^1, \theta_2^1, \theta_3^1, \theta_4^1)$ avec

θ_k^1 qui représentent les paramètres de la loi pour laquelle le traitement sera considéré comme prometteur avec une efficacité suffisante et une toxicité acceptable.

On note donc que le schéma BOP2 nécessite, au moment de toute analyse, intermédiaire ou finale, l'observation complète de la variable Y pour chaque patient inclus, c'est-à-dire l'observation d'une toxicité et/ou d'un échec ou, dans le cas contraire, le suivi de toute la durée d'observation pour chacun des 2 critères. Autrement dit, il peut être nécessaire de suspendre les inclusions dans l'attente d'une analyse intermédiaire. En règle générale, cela entraîne donc un allongement de la durée d'étude et une augmentation des coûts de la recherche.

L'estimation de la distribution de Y se fait selon une approche bayésienne. L'*a priori* conjugué à la distribution multinomiale de Y est la loi de Dirichlet de paramètres $(a_1, a_2, a_3, a_4) = c \times (\theta_1^0, \theta_2^0, \theta_3^0, \theta_4^0)$; la constante c est fixée à 1, entraînant un effective sample size $ESS = \sum_{j=1}^4 a_j = 1$, évitant ainsi de donner un poids trop important à l'*a priori* dans le cas de l'évaluation d'un traitement (9; 49; 53). Cet *a priori* représente l'information sur Y connue avant la réalisation de l'essai. Dans la majorité des cas, un *a priori* peu informatif ou non informatif est utilisé pour que les conclusions reflètent les données ($c = 1$ ici). De plus, de façon analogue à ce qui est présenté dans l'article princeps du schéma BOP, nous avons centré la distribution *a priori* sur l'hypothèse de non efficacité et de toxicité, résultant en un *a priori* qualifié de "sceptique" par Spiegelhalter (52). Si des données pré-existent au moment de la planification de l'essai, il peut être judicieux de les prendre en compte à travers cette distribution *a priori*.

Par ailleurs, la loi de Dirichlet a pour densité marginale la loi beta, ainsi $\text{Dir}(a_1, a_2) = \text{Beta}(a_1, a_2)$ et peut être agrégée de la façon suivante : si $(\alpha_1, \alpha_2, \alpha_3, \alpha_4) \sim \text{Dir}(a_1, a_2, a_3, a_4)$ alors $(\alpha_1, \alpha_2, \alpha_3 + \alpha_4) \sim \text{Dir}(a_1, a_2, a_3 + a_4)$. Grâce à ces propriétés, on peut exprimer les distributions marginales des probabilités d'efficacité p_{eff} et de toxicité p_{tox} :

$$\begin{cases} p_{\text{eff}} \sim \text{Beta}(a_1 + a_2, a_3 + a_4) \\ p_{\text{tox}} \sim \text{Beta}(a_1 + a_3, a_2 + a_4) \end{cases}$$

Si on note y_{eff} le nombre de réponses (efficacités), y_{tox} le nombre de toxicités observées lors de l'analyse intermédiaire, avec $\text{Beta}(\theta_1^0 + \theta_2^0, \theta_3^0 + \theta_4^0)$ la distribution *a priori* de la probabilité

d'efficacité et $\text{Beta}(\theta_1^0 + \theta_3^0, \theta_2^0 + \theta_4^0)$ celle de la toxicité, les probabilités *a posteriori* suivent ainsi une loi beta de la forme :

$$p_{\text{eff}}|D_n \sim \text{Beta}(\theta_1^0 + \theta_2^0 + y_{\text{eff}}, \theta_3^0 + \theta_4^0 + n - y_{\text{eff}})$$

$$p_{\text{tox}}|D_n \sim \text{Beta}(\theta_1^0 + \theta_3^0 + y_{\text{tox}}, \theta_2^0 + \theta_4^0 + n - y_{\text{tox}})$$

Sous ces hypothèses de modélisation, on définit les règles d'arrêt suivantes pour futilité et/ou toxicité :

$$\begin{cases} P(p_{\text{eff}} \leq \pi_{\text{eff}}|D_n) > C_n \\ P(p_{\text{tox}} > \pi_{\text{tox}}|D_n) > C_n \end{cases}$$

avec n le nombre de patients inclus analysés, p_{eff} la probabilité de réponse du traitement, p_{tox} la probabilité de toxicité du traitement, π_{eff} une valeur d'efficacité minimale désirable, π_{tox} une valeur de toxicité maximale tolérable, D_n les données de l'essai recueillies à l'analyse intermédiaire et C_n le seuil de décision.

C_n varie avec le nombre n de patients analysés : $C_n = 1 - \lambda(\frac{n}{N})^\gamma$. Ce seuil est optimisé par simulation (*grid search*) afin de garantir le contrôle du risque de faux positifs sous l'hypothèse d'inefficacité et de toxicité et maximiser la puissance du schéma sous l'hypothèse d'efficacité et de non toxicité.

Enfin, du fait de la monotonie de $Pr(p_{\dots} \leq \pi_{\dots}|D_n)$ avec le nombre de réponses/toxicités, il est possible de générer des tables de recrutement pour permettre aux cliniciens de monitorer facilement le recrutement des patients en fonction des réponses et toxicités observées. Ces tables indiquent le nombre maximum de réponses en dessous duquel (\leq) la règle est d'arrêter l'essai pour futilité et le nombre minimum de toxicités au dessus (\geq) duquel il est nécessaire d'arrêter l'essai pour excès de toxicité. Un exemple est présenté dans le tableau 3.

TABLE 3 – Exemple de table d’inclusion pour un schéma BOP₂ appliqué à l’évaluation jointe efficacité et toxicité dans la cohorte 1 c-MET– de l’essai METIMGAST, avec des analyses d’efficacité à $n_1 = 30$ et $n = 81$ patients et des analyses rapprochées pour la toxicité

Analyse	Nombre de patients	Stop pour futilité si $\leq \dots$ réponses *	Stop pour toxicité si $\geq \dots$ toxicités
1	5	-	4
2	10	-	6
3	15	-	7
4	20	-	9
5	30	3	11
6	40	-	14
7	50	-	16
8	60	-	19
9	70	-	21
10 (finale)	81	14	23

* Un - indique qu’il n’y a pas d’analyse d’efficacité à ce nombre de patients.

2.1.4 Le schéma TOP

Nous avons vu que le schéma BOP₂ s’appuie sur les lois beta et nécessite l’observation complète du critère évalué à chaque analyse intermédiaire. Le schéma TOP, *Time to event Bayesian Optimal Phase II design for clinical trials*, a été développé pour pallier cette contrainte face à des rythmes d’inclusions pouvant être relativement rapides, notamment dans les études multicentriques, par rapport à des fenêtres d’observation de plus en plus longues en oncologie pour les nouveaux traitements ciblés et immunothérapies (47; 54).

Brièvement, le schéma BOP₂ a été adapté pour prendre en compte les observations incomplètes pour les estimations séquentielles des distributions de Y , p_{tox} et p_{eff} , à l’aide d’une pondération de la vraisemblance selon la fraction de suivi réalisée, de façon similaire au schéma *Time-to-Event Continual Reassessment Method* (TITE-CRM) pour les essais de recherche de dose (55; 56). En pratique, les auteurs du TOP ont créé une indicatrice pour la complétude de l’observation du critère j (par exemple ici, la toxicité), chez le patient i : $\delta_{ij} = 1$ si l’observation est complète, 0 sinon. Si on note p_j la probabilité de survenue de l’évènement du critère j (dans notre exemple ici, la toxicité), les patients pour lesquels l’observation est complète (dans notre exemple : suivi complet sans toxicité ou survenue d’une toxicité) contribuent à la

vraisemblance avec $p_j^{x_{ij}} \times (1 - p_j)^{1-x_{ij}}$ avec $x_{ij} = 1$ si l'évènement j a été observé et 0 sinon. Les patients pour lesquels l'observation est incomplète, $\delta_{ij} = 0$, au moment d'une analyse intermédiaire (dans notre exemple, suivi encore en cours, sans toxicité observée) sont un mélange de patients qui n'auront pas l'évènement j et de patients qui l'auront. Leur contribution à la vraisemblance est donc : $P(\delta_{ij} = 0) = P(x_{ij} = 0) + P(T_{ij} > t_i \cap x_{ij} = 1) = 1 - P(T_{ij} \leq t_i \cap x_{ij} = 1) = 1 - P(T_{ij} \leq t_i | x_{ij} = 1) \times P(x_{ij} = 1)$ avec T_{ij} le temps jusqu'à l'évènement et t_i le temps d'observation pour le patient i . Sous l'hypothèse d'une distribution uniforme des temps d'évènements dans la fenêtre $[0; A_j]$ pour les patients ayant l'évènement, on a donc $P(T_{ij} \leq t_i | x_{ij} = 1) = \frac{t_i}{A_j} = w_{ij}$ et la contribution à la vraisemblance des patients pour lesquels $\delta_{ij} = 0$ est $1 - w_{ij}p_j \approx (1 - p_j)^{w_{ij}}$ par approximation de Taylor du 1^{er} ordre autour de 0.

Ainsi, suivant la formule de Bayes, on obtient la distribution *a posteriori* de la probabilité d'évènement j telle que :

$$\begin{aligned} P(p_j | D_n) &= \frac{P(p_j) \times P(D_n | p_j)}{P(D_n)} \\ &\propto P(p_j) \times P(D_n | p_j) \\ &\propto p_j^{a_j - 1} (1 - p_j)^{b_j - 1} \prod_{i=1}^n \left[(p_j^{x_{ij}} \times (1 - p_j)^{1-x_{ij}})^{\delta_{ij}} ((1 - p_j)^{w_{ij}})^{1-\delta_{ij}} \right] \\ &\propto p_j^{a_j - 1 + \sum_{i=1}^n (\delta_{ij} x_{ij})} (1 - p_j)^{b_j - 1 + \sum_{i=1}^n ((1-x_{ij})\delta_{ij} + w_{ij}(1-\delta_{ij}))} \end{aligned}$$

La loi *a posteriori* est donc bien de la forme d'une loi beta, avec la paramétrisation suivante :

$$\text{Beta}\left(a_j + \sum_{i=1}^n (\delta_{ij} x_{ij}); b_j + \sum_{i=1}^n ((1 - x_{ij})\delta_{ij} + w_{ij}(1 - \delta_{ij}))\right),$$

où a_j et b_j sont les hyperparamètres de la distribution *a priori* de la probabilité de l'évènement j . Les auteurs ont posé $\tilde{z}_j = \sum_{i=1}^n (\delta_{ij} x_{ij})$ qui représente le nombre d'évènements observés au temps t_i et l'*effective sample size* $ESS_j = \sum_{i=1}^n ((1 - x_{ij})\delta_{ij} + w_{ij}(1 - \delta_{ij})) = \sum_{i=1}^n (\delta_{ij} + w_{ij}(1 - \delta_{ij})) = \text{nombre de patients avec suivi complet} + \frac{\text{somme des suivis des patients en attente}}{\text{fenêtre d'observation totale de ces patients}}$.

La loi *a posteriori* s'écrit donc $P(p_j | D_n) \sim \text{Beta}(a_j + \tilde{z}_j; b_j + ESS_j - \tilde{z}_j)$.

Le schéma TOP utilise des règles d'arrêt de même expression que le schéma BOP₂ avec le

seuil C_n , calculé de la même façon, telle que décrite dans la section précédente. Il est possible, de la même manière que pour le BOP2, de générer des tables de décisions au moment de la planification de l'essai pour faciliter le recrutement et le suivi de l'essai.

2.1.5 Apport de l'information sur la toxicité issue de la cohorte c-MET+

Afin d'utiliser l'ensemble de l'information disponible, et sous l'hypothèse que le profil de toxicité du traitement évalué est indépendant du statut MET de la tumeur, le monitoring de toxicité pour l'essai METIMGAST doit prendre en compte les observations de la cohorte 2 c-MET+.

Dans le cas des schémas avec évaluation séquentielle de l'efficacité auxquels nous surajoutons une surveillance continue de la toxicité avec l'approche PP_{tox} , il est possible d'inclure les données de la cohorte 2 simplement en intégrant ces observations pour les estimations de la distribution *a posteriori* de la probabilité de toxicité. En revanche, dans le cas du schéma TOP avec critères multiples pour évaluation conjointe de l'efficacité et de la toxicité à l'aide d'un modèle multinomial, il est prévu que des analyses d'efficacité et de toxicité portent sur le même échantillon. Afin de prendre en compte l'information de la cohorte 2 dans cette situation, nous avons incorporé l'information des patients c-MET+ à travers la distribution *a priori* de la probabilité de toxicité. En effet, comme indiqué par Morita *et al.* (57), une distribution *a priori* n'est jamais que la distribution *a posteriori* issue d'un autre essai.

De plus, de façon analogue au schéma TOP, nous avons pondéré la vraisemblance de chaque observation de la cohorte 2 par son temps de suivi lorsque l'observation pour toxicité n'était pas encore complète. Ainsi, on définit un *effective sample size a priori* pour la toxicité, ESS_{tox} , comme dans la section précédente et la distribution *a priori* pour la toxicité à chaque analyse intermédiaire devient $\text{Beta}(a_{\text{tox}} + \widetilde{z}_{\text{tox}}, b_{\text{tox}} + ESS_{\text{tox}} - \widetilde{z}_{\text{tox}})$ avec $\widetilde{z}_{\text{tox}}$ le nombre de toxicités observées dans la cohorte c-MET+ au moment de l'analyse intermédiaire considérée.

Par ailleurs, il est possible de diminuer le poids des données historiques ou externes dans la distribution *a priori* (53) par un certain coefficient de pondération κ . L'hypothèse faite pour l'essai METIMGAST est que l'efficacité de la combinaison thérapeutique dépend du

statut MET de la tumeur, raison pour laquelle la cohorte c-MET+ n'est pas incluse dans l'analyse pour efficacité; mais la toxicité dans ce groupe est anticipée comme étant la même que dans la cohorte principale c-MET-. De ce fait, nous avons considéré les données c-MET+ avec un poids identique aux c-MET-, mais l'implémentation d'une pondération de la distribution *a priori* est directe et facile à intégrer. En revanche, du fait de l'actualisation de cette distribution *a priori* de la probabilité de toxicité à chaque analyse intermédiaire avec les données des observations c-MET+, la tabulation des critères d'arrêt devient difficile par le trop grand nombre de combinaisons entre les différents décomptes de toxicités possibles.

2.2 Corrélation entre efficacité et toxicité

Dans l'approche consistant à ajouter un monitoring bayésien de la toxicité (PP_{tox}) au schéma TOP utilisé pour l'efficacité seule, les événements d'efficacité et de toxicité sont considérés comme indépendants (absence de corrélation).

En réalité, en cohérence avec l'hypothèse usuelle d'une relation dose-réponse ainsi qu'une relation dose-toxicité monotones croissantes, les données suggèrent l'existence d'une corrélation positive entre l'efficacité et la toxicité des chimiothérapies cytotoxiques (58; 59). Les immunothérapies et thérapies ciblées quant à elles ont une relation dose-efficacité qui n'est pas forcément strictement monotone croissante (16). Néanmoins, il est tout de même suggéré dans la littérature une corrélation légèrement positive entre efficacité et toxicité pour les immunothérapies (60; 61).

Nous avons donc choisi d'utiliser une corrélation positive entre efficacité et toxicité pour générer les données servant au calcul des seuils de décision du schéma TOP lors de nos simulations. En effet, même si les règles d'arrêt sont dérivées à partir des probabilités marginales, le schéma TOP unifie les mesures d'efficacité et de toxicité par une unique loi multinomiale. En pratique, avec $Y \sim \text{Multinomial}(n, \theta_1, \theta_2, \theta_3, \theta_4)$, on peut obtenir la matrice de variance-covariance pour les 4 éléments de Y , (Y_1, Y_2, Y_3, Y_4) . Pour rappel, $Y_1 = \mathbb{1}(X_E = 1 \cap X_T = 1)$: "efficacité et toxicité", $Y_2 = \mathbb{1}(X_E = 1 \cap X_T = 0)$: "efficacité et pas de toxicité", $Y_3 = \mathbb{1}(X_E = 0 \cap X_T = 1)$: "pas d'efficacité et toxicité" et $Y_4 = \mathbb{1}(X_E = 0 \cap X_T = 0)$:

"pas d'efficacité et pas de toxicité" et $p_{\text{eff}} = P(Y_1 = 1) + P(Y_2 = 1)$ la probabilité d'efficacité, $p_{\text{tox}} = P(Y_1 = 1) + P(Y_3 = 1)$ la probabilité de toxicité et $p_{\text{eff} \cap \text{tox}} = P(Y_1 = 1)$ la probabilité d'avoir à la fois une réponse et une toxicité. On obtient ainsi la formule du coefficient de corrélation : $R = \frac{p_{\text{eff} \cap \text{tox}} - p_{\text{eff}} p_{\text{tox}}}{\sqrt{p_{\text{eff}}(1-p_{\text{eff}})p_{\text{tox}}(1-p_{\text{tox}})}}$. La valeur choisie pour la procédure de planification des seuils de décisions des schémas était une corrélation légèrement positive (0.214 pour l'hypothèse de non efficacité et toxicité et 0.255 pour l'hypothèse d'efficacité et non toxicité; les détails de ces calculs sont disponibles dans le manuscrit disponible en section 3.1).

Le modèle multinomial qui est utilisé par le schéma pour l'analyse séquentielle des données formalise donc naturellement la corrélation entre efficacité et toxicité. Cependant, dans la mesure où le seuil C_n est défini par simulations lors de la planification de l'essai, il nous paraît important de choisir avec soin les conditions anticipées de l'essai. Nous avons évalué par simulations la robustesse des schémas face à une discordance entre cette corrélation attendue et utilisée pour planifier l'essai et la corrélation réelle des données.

3 Article

Nous avons réalisé une étude de simulation pour évaluer les propriétés et les performances des différents schémas d'essai envisagés. Ce travail est présenté en détail dans un manuscrit pour publication dans un journal international à comité de lecture : manuscrit principal en section 3.1 et fichiers additionnels en section 3.2. Les codes R pour les simulations et analyses réalisées sont disponibles aux adresses github.com/GuillaumeMulier/multibrasBOP2 et github.com/GuillaumeMulier/MetimgastDesign.

3.1 Corps de l'article

Design of a phase II trial for digestive cancer therapeutics

Journal Title
XX(X):1-??
©The Author(s) 2022
Reprints and permission:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/ToBeAssigned
www.sagepub.com/

SAGE

Guillaume Mulier¹, Ruitao Lin², Thomas Aparicio³ and Lucie Biard¹

Abstract

Word count: 200 / 200

Introduction: New therapeutics in oncology have been challenging existing paradigms and trial designs at all phases of drug development. As a motivating example, we considered an on-going phase II trial that was planned to evaluate the combination of a MET-inhibitor and an anti-PDL1 immunotherapy to treat advanced oesogastric carcinoma. The objective was to exemplify the planning of an adaptive phase II trial with novel anti-cancer agents, including prolonged observation windows and joint sequential evaluation of efficacy and toxicity. **Methods:** We considered various candidate designs and computed decision rules assuming a slight positive correlation between efficacy and toxicity. Simulations were conducted to evaluate the operating characteristics of all designs. **Results:** Design approaches allowing continuous accrual, such as the time-to-event Bayesian Optimal Phase II design (TOP) showed good operating characteristics while ensuring reduced trial duration. Designs relying on a separate posterior probability monitoring of toxicity were less powerful but were less sensitive to mis-specification of the correlation between efficacy and toxicity. **Discussion:** While specification of design working hypotheses requires caution, Bayesian approaches such as the TOP design had desirable operating characteristics and allowed incorporating external information, such as toxicity data from an on-going mutational cohort, via *a priori* distributions in estimation procedures.

Keywords

Phase II, Adaptive design, Bayesian, Oncology

Introduction

The development of drugs in oncology has long driven statistical innovations for trial designs^{1,2}. The severity of the disease has led to consider drugs with non negligible toxicity^{3,4}. This particular benefit/risk ratio result in trials enrolling patients, from the early phases of development, rather than healthy volunteers⁵. Methodological challenges have arisen over the past decades with new types of treatments such as immunotherapy⁶. Their mode

of action differs from that of conventional cytotoxic chemotherapy, resulting in different kinetics in both toxicity

¹INSERM UMR 1153, Saint-Louis hospital, France

²Department of Biostatistics, MD Anderson Cancer Center, USA

³Hépto-gastro-entérologie, Saint-Louis hospital, France

Corresponding author:

Guillaume Mulier, INSERM UMR 1153, Hôpital Saint-Louis, 1 avenue Claude Vellefaux, 75010 Paris, France
Email: dry12@hotmail.fr

and efficacy, often with prolonged timeframes. Such drugs are administered in a continuous manner, as opposed to intermittent cycles for cytotoxic chemotherapy.

Specifically, from a trial planning perspective, immunotherapy usually implies long-term endpoints and potentially non-monotone dose-activity relationships^{1,7}. These particularities challenged the existing designs calibrated to evaluate cytotoxic chemotherapies^{6,8-10}.

Phase II trials, notably in oncology, aim at identifying promising therapies while ruling out the unpromising as soon as possible, prior to large-scale phase III studies¹¹. Adaptive designs are of choice allowing early stopping rules in trials¹². Moreover, sample sizes are often limited and designs may include the joint evaluation of efficacy and toxicity to some extent. Overall, all these particularities may result in the need for complex designs and multiple analyses may be combined with caution to avoid an inflated risk of false positive¹³⁻¹⁵.

In this paper, we present the statistical planning of a phase II non comparative trial using an innovating tailored design, for the evaluation of a novel therapeutic drugs combination for advanced oesogastric cancer.

In the following sections, we first present the trial and its clinical background. We then consider several design options given the specific clinical settings and compare their operating characteristics in a simulation study. Lastly we provide points for discussion.

The METIMGAST trial

Oesogastric adenocarcinoma is part of gastric cancers which are the 3rd most deadly cancer worldwide¹⁶. Oesogastric adenocarcinoma has a serious prognosis even after first-line treatment by chemotherapy and/or surgery^{17,18}. 2nd-line treatments still come with low survival¹⁹⁻²¹. Some preclinical evidences suggests synergistic action of the association of association MET inhibitor and anti-PDL1^{22,23}. The association has been studied in non small cell lung carcinoma (NCT03647488) and advanced hepatocellular carcinoma

(NCT02795429); and is currently evaluated in melanoma (NCT03484923) and breast cancer (NCT03742349).

METIMGAST trial is an investigator-initiated multicenter phase II trial across France, to evaluate the combination of capmatinib and spartalizumab in advanced oesogastric adenocarcinoma in adults (NCT05135845).

A total of 90 patients were planned to be recruited. Since MET amplification may be predictive of the drug efficacy, and given the low prevalence of MET amplification, we planned a separate evaluation of efficacy in patients without MET amplification. Nevertheless, for the toxicity evaluation, we planned to incorporate information from all patients, regardless of MET amplification status. Assuming a 10%-prevalence of MET amplification, we expect about 81 patients without MET amplification (Group 1), and 9 patients with MET amplification (Group 2).

The primary endpoint for efficacy was the overall response rate (ORR) defined as the proportion of patients with partial or complete remission according to RECIST v1.1 criteria within 6 months after inclusion (that is 8 cycles of treatment). The main toxicity endpoint was the occurrence of an unacceptable toxicity event within 42 days after inclusion (corresponding to 2 cycles of treatment), defined using NCI-CTCAE v5 criteria as at least one of:

- Adverse event of grade ≥ 3 at least possibly related to the treatment or unrelated to disease, disease progression, intercurrent illness, or concomitant medications;
- Non-hematological toxicity \geq grade 3 (except for nausea, vomiting, fatigue);
- Recurring grade 2 pneumonitis;
- Grade ≥ 2 myocarditis;
- Autoimmune hemolytic anemia, hemolytic uremic syndrome, or acquired hemophilia grade ≥ 3 ;
- Guillain-Barre, severe peripheral or autonomic neuropathy, or transverse myelitis;

- Encephalitis or aseptic meningitis;
- Laboratory abnormality \geq grade 3 lasting more than 7 consecutive days (except grade 3/4 nephritis, combined elevation of total bilirubin and AST/ALT, hyperglycaemia and changes in serum electrolytes/enzymes without clinical impact);
- Hematological toxicities;
- Severe adverse event;
- Significant drug related toxicity.

Thus the observation windows were 180 days for efficacy and 42 days for toxicity. The anticipated accrual rate was 5 patients per month, that is a mean interpatient time of 6 days.

Methods

The planned design

Considering the anticipated accrual rate relatively to the efficacy observation window, the trial was initially planned using an adaptive Bayesian phase 2 design: the Time-to-event Optimal Phase 2 design²⁴, for the efficacy endpoint in Group 1 with analyses at 30 and 81 patients. The TOP design is a Bayesian framework based on BOP-design²⁵ allowing to assess multiple endpoints at the same time though the use of multinomial distribution. It takes into account the information of pending patients via a weighting in the likelihood, allowing to shorten the duration of trials and gain in power compared with the BOP approach. A simple estimation of ORR was planned in Group 2, without interim analysis. Motivated by safety concerns, a toxicity monitoring was added with more frequent looks (at 5, 10, 15, 20, 30, 40, 50, 60, 70, 80 and 90 patients over both groups). This toxicity monitoring was planned using a posterior distribution approach^{26,27} with the following decision rule: $Pr(p_{\text{tox}} > 0.25|D_n) > 0.95$, where p_{tox} is the true probability of unacceptable toxicity. Working hypotheses on efficacy were the following: $H_0 : p_{\text{eff}} \leq 0.15$ and $H_1 : p_{\text{eff}} = 0.30$.

The TOP design was parameterized to ensure 90% power under these hypotheses, given a 5% type I error rate. The design's parameters were $\lambda = 0.92/\gamma = 0.97$, which provided 94.76% power under H_1 . By simulations, it was assessed that the addition of the proposed toxicity monitoring would keep power above 90% when toxicity risk is low ($p_{\text{tox}} \leq 0.20$) and that in the case of unacceptable toxicity the trial will be stopped in 49% of cases when $p_{\text{tox}} = 0.30$ and 95% when $p_{\text{tox}} = 0.40$. Retrospectively, to ensure lower proportion of false positive trials when $p_{\text{tox}} = 0.30$ we changed the decision rule to $Pr(p_{\text{tox}} > 0.20|D_n) > 0.95$. That decision rule was evaluated in combination with the planned TOP-design and led to a type I error rate of 0.6% and a power of 82.7% (2.6% and 92.5% respectively for the initial design). This loss in power was justified by the will to limit the false positive rate when the treatment is toxic and was evaluated through simulation (see additional file 1).

Studied designs

Overall we compared the main proposal described above (denoted "TOP_{eff}+PP_{tox}" hereafter) with five other approaches adapted to our clinical setting in terms of trial design. We evaluated designs omitting the time-to-event information on the outcomes using two designs with strictly binary endpoints, and three approaches relying on variations of the TOP design. For all approaches, sample sizes (interim and final) for efficacy analyses were kept identical to the main proposal: 30 and 81 patients. Overall, six approaches were evaluated:

1. TOP design to assess the efficacy in group 1, associated with a parallel toxicity monitoring based a posterior probability rule at 5/10/15/20/30/40/50/60/70/80/90 patients in both groups ("TOP_{eff}+PP_{tox}").
2. Simon's 2 stage design²⁸ to assess the efficacy in group 1, associated with a parallel toxicity monitoring based a posterior probability rule

- at 5/10/15/20/30/40/50/60/70/80/90 patients in both groups (referred as "Simon + PP_{tox}" in the following).
3. BOP2 design for efficacy²⁵ with interim analyses at 30 and 81 patients in Group 1 and the posterior probability's approach for toxicity assessment, at the same numbers of patients as "TOP_{eff}+PP_{tox}" in both groups (denoted "BOP_{eff}+PP_{tox}" hereafter);
 4. TOP design with two endpoints, efficacy and toxicity, with analyses at 30 and 81 patients for both endpoints, using only Group 1 (referred as "TOP_{eff/tox} with 2 analyses");
 5. TOP design with two endpoints, efficacy and toxicity, with analyses at 30 and 81 patients for efficacy and at (5/10/15/20/30/40/50/60/70/81) patients for toxicity, using only Group 1 (denoted "TOP_{eff/tox} with close monitoring of toxicity" below);
 6. TOP design with two endpoints, efficacy and toxicity, with analyses at 30 and 81 patients for efficacy and at (5/10/15/20/30/40/50/60/70/81) patients for toxicity, in Group 1, but incorporating accumulated data from Group 2, as it becomes available for the assessment of toxicity. The information from Group 2 is incorporated via the prior distribution for the probability of toxicity (referred as "TOP_{eff/tox} with update of the prior of toxicity" hereafter).

Design calibration

All designs were calibrated based on the following working hypotheses, on efficacy: $p_{0,eff} = 0.15$ and $p_{1,eff} = 0.30$, consistently with H_0 and H_1 ; and similarly on toxicity, $p_{0,tox} = 0.30$ and $p_{1,tox} = 0.20$. The maximum consented risk of erroneous conclusion to a promising treatment (efficacious and non toxic) was set to 5% under the working hypothesis $\{p_{0,eff} = 0.15; p_{0,tox} = 0.30\}$.

Decision boundaries for the "TOP_{eff}+PP_{tox}" and "BOP_{eff}+PP_{tox}" were tuned following the BOP2/TOP procedure for efficacy^{24,25} and using the posterior

distribution on toxicity risk^{26,27} for toxicity, under the above listed hypotheses on efficacy and toxicity; parameters for the BOP2/TOP efficacy decision boundaries were $\lambda = 0.92/\gamma = 0.97$ (using notations from the original TOP paper).

The designs "TOP_{eff/tox} with close monitoring of toxicity" and "TOP_{eff/tox} with update of the prior of toxicity" used the same decision boundaries formula. Calculation of λ and γ was adapted in additional file 1 and we obtained: $\lambda = 0.69/\gamma = 0.98$, which were optimized under the hypothesis of a positive correlation between efficacy and toxicity. They corresponded to a type I error rate of 3.93% and a planned power of 89.71%.

The design "TOP_{eff/tox} with 2 analyses" used another set of parameters: $\lambda = 0.69/\gamma = 0.68$, which were optimized under the hypothesis of a positive correlation between efficacy and toxicity to have a planned type I error rate of 4.75% and a planned power of 92.36%.

Lastly, for comparative purposes, we also implemented a Simon's 2 stage design minimizing the average sample size under $p_{0,eff} = 0.15$ with the first analysis at 30 patients and the final at 81 patients. This design calibration is described in additional file 1. Decision boundaries in terms of efficacy and toxicity event counts are available in Additional file 2 for all designs.

Of note, in the case of combining TOP and posterior probabilities' approaches, it is not straightforward to take into account the correlation between efficacy and toxicity when calibrating the design. However, when a TOP design combining efficacy and toxicity endpoints is used, the impact of the correlation between efficacy and toxicity is accounted for with the use of a multinomial sampling distribution. The pair of calibration parameters (γ, λ) for the BOP2/TOP approaches may vary depending on the assumed correlation between efficacy and toxicity. In the original trial design, efficacy and toxicity were considered independent, but a

probable hypothesis would be that efficacy and toxicity are positively correlated^{29,30}.

We conducted a few simulations to assess how the correlation between efficacy and toxicity might impact the calibration of the TOP_{eff/tox} design. Figures S4 and S5 in additional file 1 illustrates the variation of the different design parameters: λ , γ , planned type I error rate (denoted Planned α) and power (denoted Planned power), according to the correlation between efficacy and toxicity, under $\{p_{0,\text{eff}} = 0.15; p_{0,\text{tox}} = 0.30\}$ and $\{p_{1,\text{eff}} = 0.30; p_{1,\text{tox}} = 0.20\}$. We found that the design's parameters varied with the degree of correlation between efficacy and toxicity under $\{p_{0,\text{eff}} = 0.15; p_{0,\text{tox}} = 0.30\}$: while there was no clear pattern for type I error rate, the power increased with the correlation under $\{p_{0,\text{eff}} = 0.15; p_{0,\text{tox}} = 0.30\}$. As expected, the correlation under $\{p_{1,\text{eff}} = 0.30; p_{1,\text{tox}} = 0.20\}$ had less influence on the design's calibration and operating characteristics, with only slight changes in the expected power.

With more frequent analyses for toxicity, we found a similar pattern for the design's parameters values, dependent on the correlation between efficacy and toxicity under $\{p_{0,\text{eff}} = 0.15; p_{0,\text{tox}} = 0.30\}$. However, planned power slightly decreased compared to the design with only 2 analyses for toxicity (additional file 1).

Simulation settings

We evaluated the operating characteristics of the five designs under 10 scenarios of true efficacy, p_{eff} , and toxicity, p_{tox} , as reported in table 1. For each scenario, correlation variations were specified, exploring a range of 6 values of correlation between efficacy and toxicity, from R_{\min} to R_{\max} , defined as follows:

- $R_{\min} : Pr(\text{Eff} \cap \text{Tox}) = \max(0, Pr(\text{Eff}) + Pr(\text{Tox}) - 1)$;
- $R_{\text{neg}} = \frac{R_{\min}}{2}$;
- $R_{\text{ind}} = 0$;
- $R_{\text{pos},1} = \frac{R_{\max}}{3}$;

- $R_{\text{pos},2} = \frac{2R_{\max}}{3}$;
- $R_{\max} : Pr(\text{Eff} \cap \text{Tox}) = \min(Pr(\text{Eff}), Pr(\text{Tox}))$.

Table 1 presents the 10 scenarios with their respective true probabilities of efficacy and toxicity. Given these probabilities, the minimum and maximum correlation between efficacy and toxicity, R_{\min} and R_{\max} respectively, were computed, as defined above, and are reported in table 1. Remaining correlation variations that were evaluated can be derived using the above formulas. For example, scenario 1 with correlation $R_{\text{pos},1}$ corresponds to $\{Pr(\text{Eff} \cap \text{Tox}) = 0.08, Pr(\text{Eff} \cap \overline{\text{Tox}}) = 0.07, Pr(\overline{\text{Eff}} \cap \text{Tox}) = 0.22, Pr(\overline{\text{Eff}} \cap \overline{\text{Tox}}) = 0.63\}$ and scenario 2 to $\{Pr(\text{Eff} \cap \text{Tox}) = 0.11, Pr(\text{Eff} \cap \overline{\text{Tox}}) = 0.19, Pr(\overline{\text{Eff}} \cap \text{Tox}) = 0.09, Pr(\overline{\text{Eff}} \cap \overline{\text{Tox}}) = 0.61\}$. For the sake of simplicity, results with a positive correlation between efficacy and toxicity ($R_{\text{pos},1}$) are primarily reported in the following section. We also presented the results of the 6 correlations presented above applied on scenarios 1 and 2 in order to assess the risk of false positive conclusions under the inefficacy and toxicity case, and the risk of false negative under the case of desirable efficacy and unacceptable toxicity. Additional results on the other scenarios are available in additional file 3.

For each scenario, 10000 simulated trials were generated. We estimated the following characteristics for each candidate design: probability of conclusion of efficacy and no toxicity (positive trial), average trial duration, average included sample size, probability of early stopping. All analyses were performed on R statistical software version 4.0.2.

Results

Operating characteristics

Figure 1 represents the percentage of conclusion to a promising treatment (efficacy and no toxicity) across the 10 scenarios with a positive correlation between efficacy

and toxicity (ranging from 0.21 to 0.33 depending on the scenario, see table 1 for details).

Results from scenario 1 showed that the risk of false positive is controlled under $\{p_{0,\text{eff}} = 0.15; p_{0,\text{tox}} = 0.30\}$, whichever the design, with posterior probabilities' approaches being the most conservative. Scenario 2 allowed assessing the power under the anticipated $\{p_{1,\text{eff}} = 0.30; p_{1,\text{tox}} = 0.20\}$: power was the greatest using the TOP_{eff/tox} approaches: "TOP_{eff/tox} with 2 analyses" (90.8%), "TOP_{eff/tox} with update of the prior of toxicity" (88.1%) and "TOP_{eff/tox} with close monitoring of toxicity" (87.9%) approaches. The remaining designs had lower power (82.8% for "TOP_{eff} + PP_{tox}"; 74.0% for "BOP_{eff} + PP_{tox}" and 72.8% "Simon + PP_{tox}"). Simon's design seemed to have similar percentage of acceptance of the drug as BOP2 design. Among strategies with TOP-design for both efficacy and toxicity endpoints, more frequent interim analyses resulted in a more conservative design as seen across all 10 scenarios. In general, in the setting that we studied, the approach using the TOP for efficacy and toxicity is less conservative than the approaches with the posterior probabilities design used for toxicity monitoring. Of note, when the treatment is not effective, the lower the toxicity, the higher the risk of false positive become, particularly for TOP-only approaches.

Last, the approach including the update of the toxicity prior using data from Patient group 2 allowed to slightly better control the risk of false positive compared to the other TOP-only strategies.

The average number of patients per trial, the proportion of early stopping and the mean duration of a trial with a slight positive correlation between efficacy and toxicity are represented in additional file 3 (figures S6 to S8).

Overall, the approaches "BOP_{eff} + PP_{tox}" and "Simon + PP_{tox}" resulted in higher proportion of early stopping whichever the scenario. Consistently, the sample size was smaller with these designs.

Regarding the remaining 4 designs, the update of the prior and the "BOP_{eff}+PP_{tox}" strategies tend to have a greater proportion of early stopping, consistent with a reduced risk of false positives and a reduced sample size.

The average duration of a trial was greater with the "BOP_{eff}+PP_{tox}" design, which was expected given that the BOP2 design requires to wait for complete observations of all included patients at each interim analysis. The same holds for "Simon + PP_{tox}" design. For the remaining 4 designs, the duration was consistent with the mean number of patients per trial.

Correlation between efficacy and toxicity

We then assessed if the approaches were robust to discrepancies between planning hypothesis and reality in terms of correlation between efficacy and toxicity. In other words, the designs' parameters were calculated using a positive correlation between efficacy and toxicity whereas the true correlation between efficacy and toxicity in the trial is not always positive.

For illustrative purposes, we first evaluated the type I error rate and power of the proposed design under scenario 1 and 2, with correlation between efficacy and toxicity. The correlation is linearly associated with $p_{\text{eff} \cap \text{tox}}$, so we made vary $p_{\text{eff} \cap \text{tox}}$ between its minimum ($\max(0, p_{\text{eff}} + p_{\text{tox}} - 1)$) and its maximum ($\min(p_{\text{eff}}, p_{\text{tox}})$). Results are presented in the figure 2.

Even though The TOP design directly accounts for the correlation between efficacy and toxicity when planning the trial using a multinomial sampling distribution. Its type I error rate was the most impacted by misspecified correlation as compared to designs with posterior probabilities' approaches. Specifically, for all designs, type I error rate increased when correlation was smaller, but to a lesser extent with the 3 posterior probabilities' designs. Power was about constant for all designs, with the TOP-only strategies having the greater power.

We then report the results under scenarios 1 to 10, with a varying correlation as described above, while the parameters were still optimized for a slight positive correlation (corresponding to $R_{\text{pos},1}$). These results are available in additional file 3 (figures S9 to S12).

Scenario 1 shows again the augmentation of risk of false positive under the null hypothesis when the correlation between efficacy and toxicity decreases. And that TOP exclusive approaches are more affected than the posterior probabilities' approaches. In general, for all designs, when the correlation between efficacy and toxicity increases, the proportion of positive trials decreases in all scenarios except scenario 2. The effect of the correlation between efficacy and toxicity on the mean number of patients recruited is very limited and there are not many differences. The same applies for the proportion of premature stopping and mean duration of a trial. The overall conclusions are consistent with what was depicted for the positive correlation.

Discussion

We report the comparative study of several approaches of trial design, for a phase II clinical evaluations. More specifically, our work was motivated by a phase II clinical trial evaluating a new combination of drugs in the treatment of oesogastric cancer. The joint evaluation of efficacy and safety was required due to the non negligible toxicity risk of these anti-cancer agents.^{31,32} We assessed the operating characteristics of candidate designs and their robustness to departure from planned conditions and hypotheses.

Namely, we compared designs using standard binomial modelling of the study outcomes, either using Simon's optimal design²⁸ or the Bayesian optimal design (BOP2 design)²⁵, combined with a Bayesian posterior probability approaches for toxicity monitoring^{26,27}, to designs incorporating time-to-event information on the outcomes, via a weighted likelihood, using the TOP design²⁴.

Overall, the TOP-design approaches showed greater power while controlling the type I error rate under the specified conditions prior to the trial. When considering toxicity monitoring, although the posterior probability approach resulted in a decreased power, its control of type I error rate was stricter and less sensitive to misspecification of the correlation between efficacy and toxicity. Conversely, using the TOP design for both efficacy and toxicity endpoints, an overestimated correlation between efficacy and toxicity at the planning stage can lead to higher risk of false positive. Lastly, the TOP strategies with toxicity analyses more frequent than efficacy ones were less powerful, and the addition of external information for toxicity evaluation allowed decreased proportion of false positive trials in case of a toxic treatment.

The TOP approaches represent an advantageous choice in clinical settings requiring long-term endpoints, which is frequent in oncology^{6,33,34}. Subsequent prolonged observation windows^{8,10} can be challenging in designing early phase trials, in particular in case of interim analyses. A major concern is then to determine when interim analyses will be performed. In the case of a design relying on strictly binary data for endpoints such a Simon's optimal design, complete observations must be obtained to perform any analysis, potentially resulting in a waiting period with suspended inclusions over the course of the trial. The TOP design allows accounting for any available data at the time of an analysis, even incomplete observations, incorporating the fraction of available follow-up in the model via a weighted likelihood. Straightforwardly, the trial may allow continuous patient enrollment and the trial duration can be reduced compared to a standard design. Moreover, in our setting, we found that the TOP approach was more powerful than the BOP, at the cost of a slight increase of the risk of false positives.

Concerning the observation window, safety monitoring approaches with the posterior probability of toxicity relied on binary data and therefore used complete toxicity

observations only at the time of the enrollment of the next patient. In our settings, the observation window was shorter for toxicity (42 days) than for efficacy (6 months), while the anticipated average interpatient time was 6 days, for accrual. Should the toxicity observation window be longer and/or farther from the range of the interpatient time, given the default assumption of a uniform distribution of the toxicities over time, adapting the posterior probability approach with a time-to-event component, rather than excluding incomplete observations from interim analyses, should be considered to avoid false positive trials in case of toxic drugs. Indeed, the distribution of the time to toxicity may also affect the designs' operating characteristics.

Furthermore, the decision thresholds for futility stopping were lower when the TOP was optimized only for efficacy (separate posterior probability monitoring for toxicity) than for joint efficacy and toxicity outcomes. Consequently, the posterior probabilities' approach had higher chances of stopping than the TOP approaches with joint outcomes. In case of a treatment that needs to be very closely evaluated for toxicity, a joint TOP approach should be preferred. Last, it has been shown that in the case of frequentist approaches with efficacy stopping rule, more frequent looks can sometimes lead to an increased proportion of positive trials and thus to an overestimation of the efficacy for example³⁵. When applied to a toxicity outcome, this translates in a risk of more conservative trials, the more frequent are interim analyses.

Conclusion

In our setting, we found that a design combining the TOP design for efficacy and a posterior probability monitoring on toxicity results in a greater power while controlling the risk of false positive. In case extra caution is needed given a drug toxicity profile, a TOP design, with joint efficacy and toxicity outcomes, ensures a more conservative approach.

Furthermore, an advantage of the joint TOP approach for efficacy and toxicity is that the number of patients and the duration of the trial are lower than with posterior probabilities' approach, all the more with prolonged toxicity observation windows.

Also, we found that external data on toxicity, corresponding to a low-prevalence mutational subgroup in our setting, could be incorporated, in the case of Bayesian approaches through the toxicity prior, resulting in more stringent boundaries for safety assessment.

References

1. Li Q, Lin J and Lin Y. Adaptive design implementation in confirmatory trials: methods, practical considerations and case studies. *Contemporary Clinical Trials* 2020; 98: 106096.
2. Day D and Siu LL. Approaches to modernize the combination drug development paradigm. *Genome medicine* 2016; 8(1): 1–14.
3. Markman TM and Markman M. Cardio-oncology: mechanisms of cardiovascular toxicity. *F1000Research* 2018; 7.
4. Stone JB and DeAngelis LM. Cancer-treatment-induced neurotoxicity—focus on newer treatments. *Nature reviews Clinical oncology* 2016; 13(2): 92–105.
5. Karakunnel JJ, Bui N, Palaniappan L et al. Reviewing the role of healthy volunteer studies in drug development. *Journal of translational medicine* 2018; 16(1): 1–15.
6. Postel-Vinay S, Aspeslagh S, Lanoy E et al. Challenges of phase 1 clinical trials evaluating immune checkpoint-targeted antibodies. *Annals of Oncology* 2016; 27(2): 214–224.
7. Mozgunov P and Jaki T. An information theoretic phase i–ii design for molecularly targeted agents that does not require an assumption of monotonicity. *Journal of the Royal Statistical Society Series C, Applied Statistics* 2019; 68(2): 347.
8. Ghisoni E, Wicky A, Bouchaab H et al. Late-onset and long-lasting immune-related adverse events from immune checkpoint-inhibitors: An overlooked aspect in immunotherapy. *European journal of cancer* 2021; 149: 153–164.
9. Tang J, Shalabi A and Hubbard-Lucey VM. Comprehensive analysis of the clinical immuno-oncology landscape. *Annals of Oncology* 2018; 29(1): 84–91.
10. Song G and Ivanova A. Enrollment and stopping rules for managing toxicity requiring long follow-up in phase ii oncology trials. *Journal of biopharmaceutical statistics* 2015; 25(6): 1206–1214.
11. Gray R, Manola J, Saxman S et al. Phase ii clinical trial design: methods in translational research from the genitourinary committee at the eastern cooperative oncology group. *Clinical cancer research* 2006; 12(7): 1966–1969.
12. Li Q, Lin J, Liu M et al. Utilizing surrogate endpoints in adaptive designs with delayed treatment effect. *Statistics in Biopharmaceutical Research* 2021; (just-accepted): 1–15.
13. Ryan EG, Brock K, Gates S et al. Do we need to adjust for interim analyses in a bayesian adaptive trial design? *BMC medical research methodology* 2020; 20(1): 1–9.

14. Stallard N, Todd S, Ryan EG et al. Comparison of bayesian and frequentist group-sequential clinical trial designs. *BMC medical research methodology* 2020; 20(1): 1–14.
15. Jennison C and Turnbull BW. *Group sequential methods with applications to clinical trials*. CRC Press, 1999.
16. Bray F, Ferlay J, Soerjomataram I et al. Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries (vol 68, pg 394, 2018). *Ca-a Cancer Journal for Clinicians* 2020; 70(4): 313–313.
17. Lemoine N, Adenis A, Bouche O et al. Signet ring cells and efficacy of first-line chemotherapy in advanced gastric or oesogastric junction adenocarcinoma. *Anticancer research* 2016; 36(10): 5543–5549.
18. Al-Batran SE, Homann N, Pauligk C et al. Perioperative chemotherapy with fluorouracil plus leucovorin, oxaliplatin, and docetaxel versus fluorouracil or capecitabine plus cisplatin and epirubicin for locally advanced, resectable gastric or gastro-oesophageal junction adenocarcinoma (flot4): a randomised, phase 2/3 trial. *The Lancet* 2019; 393(10184): 1948–1957.
19. Thuss-Patience PC, Kretschmar A, Bichev D et al. Survival advantage for irinotecan versus best supportive care as second-line chemotherapy in gastric cancer—a randomised phase iii study of the arbeitsgemeinschaft internistische onkologie (aio). *European journal of cancer* 2011; 47(15): 2306–2314.
20. Kang JH, Lee SI, Lim D et al. Salvage chemotherapy for pretreated gastric cancer: a randomized phase iii trial comparing chemotherapy plus best supportive care with best supportive care alone. *Clin Oncol* 2012; 30(13): 1513–1518.
21. Ford HE, Marshall A, Bridgewater JA et al. Docetaxel versus active symptom control for refractory oesophagogastric adenocarcinoma (cougar-02): an open-label, phase 3 randomised controlled trial. *The lancet oncology* 2014; 15(1): 78–86.
22. Li H, Li CW, Li X et al. Met inhibitors promote liver tumor evasion of the immune response by stabilizing pdl1. *Gastroenterology* 2019; 156(6): 1849–1861.
23. Glodde N, Bald T, van den Boorn-Konijnenberg D et al. Reactive neutrophil responses dependent on the receptor tyrosine kinase c-met limit cancer immunotherapy. *Immunity* 2017; 47(4): 789–802.
24. Lin R, Coleman RL and Yuan Y. Top: Time-to-event bayesian optimal phase ii trial design for cancer immunotherapy. *JNCI: Journal of the National Cancer Institute* 2020; 112(1): 38–45.
25. Zhou H, Lee JJ and Yuan Y. Bop2: Bayesian optimal design for phase ii clinical trials with simple and complex endpoints. *Statistics in medicine* 2017; 36(21): 3302–3314.
26. Ivanova A, Song G, Marchenko O et al. Monitoring rules for toxicity in phase ii oncology trials. *Clinical Investigation* 2015; 5(4): 373–381.
27. Thall PF, Simon RM and Estey EH. New statistical strategy for monitoring safety and efficacy in single-arm clinical trials. *Journal of Clinical Oncology* 1996; 14(1): 296–303.
28. Simon R. Optimal two-stage designs for phase ii clinical trials. *Controlled clinical trials* 1989; 10(1): 1–10.
29. Palmieri DJ and Carlini MS. Immune checkpoint inhibitor toxicity. *Current oncology reports* 2018; 20(9): 1–12.
30. Das S and Johnson DB. Immune-related adverse events and anti-tumor efficacy of immune checkpoint inhibitors. *Journal for immunotherapy of cancer* 2019; 7(1): 1–11.
31. Bryant J and Day R. Incorporating toxicity considerations into the design of two-stage phase ii clinical trials. *Biometrics* 1995; : 1372–1383.
32. Tournoux C, De Rycke Y, Médioni J et al. Methods of joint evaluation of efficacy and toxicity in phase ii clinical trials. *Contemporary clinical trials* 2007; 28(4): 514–524.
33. Koopmeiners JS and Modiano J. A bayesian adaptive phase i–ii clinical trial for evaluating efficacy and toxicity with delayed outcomes. *Clinical Trials* 2014; 11(1): 38–48.
34. Kaufman HL, Atkins MB, Subedi P et al. The promise of immuno-oncology: implications for defining the value of cancer treatment. *Journal for ImmunoTherapy of Cancer* 2019; 7(1): 1–11.
35. Bassler D, Montori VM, Briel M et al. Early stopping of randomized clinical trials for overt efficacy is problematic. *Journal of clinical epidemiology* 2008; 61(3): 241–246.

Tables

Table 1. Simulation scenarios, with $Pr(\text{Eff})$ the true probability of efficacy, $Pr(\text{Tox})$ the true probability of toxicity, R_{\min} and R_{\max} the true minimum and maximum correlations respectively, given the event probabilities.

Scenario	Description	$Pr(\text{Eff})$	$Pr(\text{Tox})$	R_{\min}	R_{\max}
1	H_0	0.15	0.30	-0.27	0.64
2	H_1	0.30	0.20	-0.33	0.76
3	Intermediate	0.20	0.25	-0.29	0.87
4	Intermediate (2)	0.25	0.25	-0.33	1.00
5	Inefficacious	0.15	0.20	-0.21	0.84
6	Inefficacious (2)	0.10	0.15	-0.14	0.79
7	Intermediate efficacy	0.20	0.20	-0.25	1.00
8	Intermediate efficacy (2)	0.20	0.30	-0.33	0.76
9	Toxic	0.30	0.30	-0.43	1.00
10	Toxic (2)	0.40	0.35	-0.60	0.90

Figures

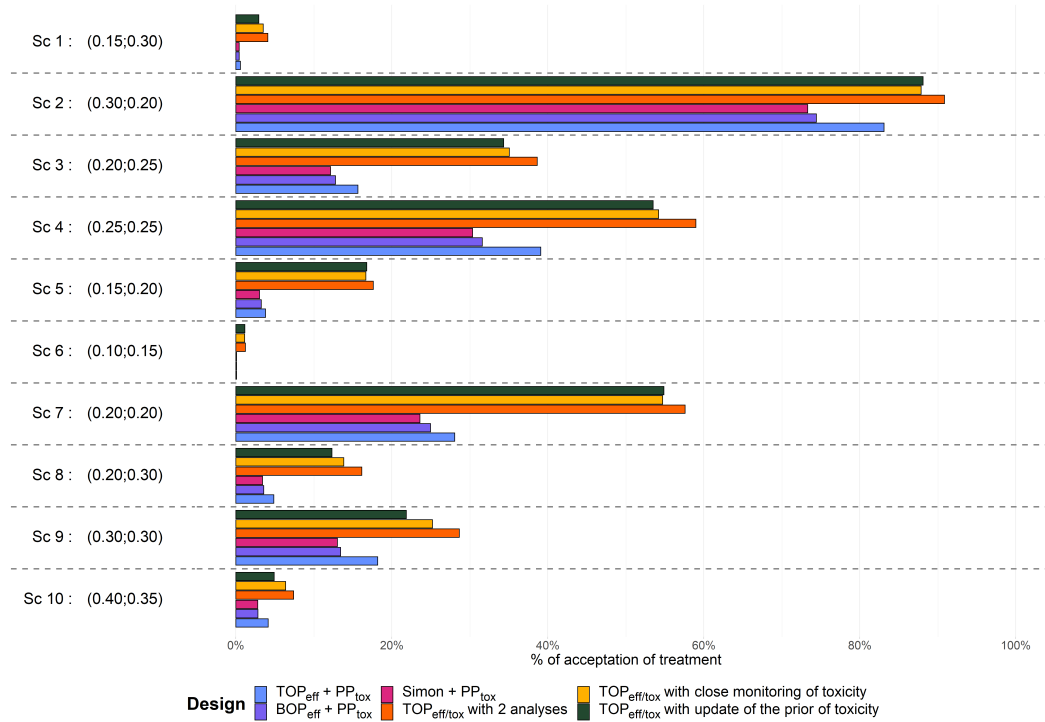


Figure 1. Percentage of conclusion of efficacy and no toxicity in the 10 scenarios with a positive correlation between efficacy and toxicity. Numbers in parentheses represent p_{eff} and p_{tox} for each scenario.

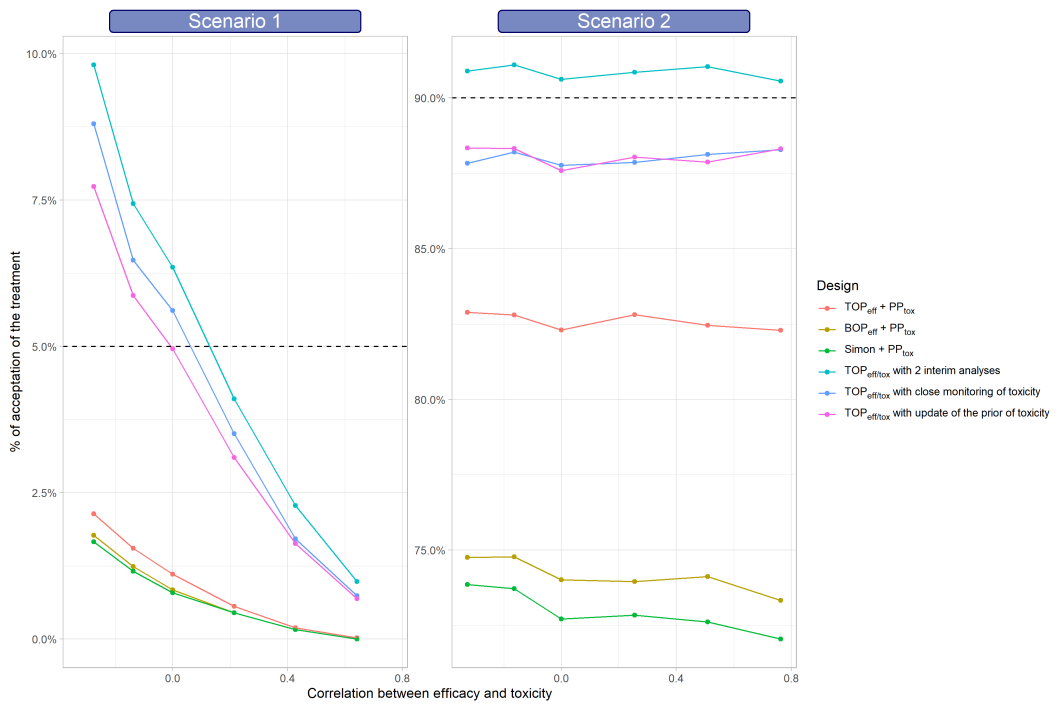


Figure 2. Percentage of conclusion of efficacy and no toxicity relative to the correlation between efficacy and toxicity in the setting. Left panel: type I error rate; Right panel: power. (Scale is panel-specific.)

3.2 Fichiers additionnels

3.2.1 Additional file 1 – Calibration of parameters for each design

Calibration of PP_{tox} procedure

We considered two calibration approaches for the PP_{tox} toxicity monitoring : either considering only toxicity data (omitting the parallel efficacy analysis) or accounting for the joint evaluation of toxicity with efficacy.

Considering only toxicity analyses

We computed the type I error rate (proportion of rejecting the toxicity hypothesis when the treatment is toxic) and power (proportion of rejecting the toxicity hypothesis when the treatment is truly non toxic) for a set of values for π_{tox} and τ . As a reminder, the decision rule for toxicity monitoring was to terminate the trial for toxicity if the posterior probability that the toxicity rate p_{tox} being above a threshold π_{tox} was greater than τ , $0 < \tau < 1$ (e.g. $Pr(p_{\text{tox}} > 0.25|D_n) > 0.95$).

We generated the data of 10000 simulated trials with 90 patients each, sampling toxicity observations from a binomial distribution (either with true toxicity probability $p_{\text{tox}} = 0.30$ (H_0) or $p_{\text{tox}} = 0.20$ (H_1)). Values of π_{tox} were explored between 0.20 and 0.30, and values of τ between 0.86 and 0.99. The results are presented in figure S1.

Power and type I error rate for PP_{tox} design on toxicity

Stopping rule : $P(p_{tox} > \pi_{tox} | D_n) > \tau$

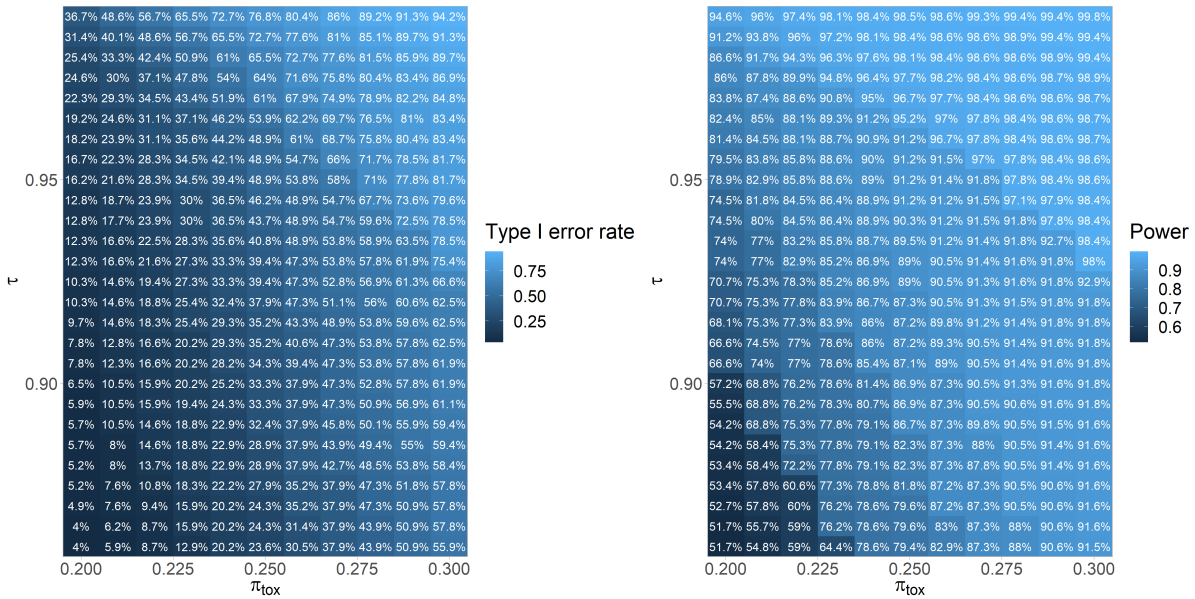


FIGURE S1 – Type I error rate and power on toxicity assessment, relative to parameters π_{tox} and τ in the decision rule

As expected, the greater π_{tox} , the less trials were stopped for toxicity because $Pr(p_{tox} > \pi_{tox} | D_n)$ is monotonically decreasing with π_{tox} . Similarly, the greater τ , the less frequent early stoppings for toxicity. With $\pi_{tox} = 0.25$ and $\tau = 0.95$ as proposed in the original METIMGAST proposition, there was 91.2% of trials concluding to non toxicity when the toxicity is truly at 0.20, and conversely 48.9% of the simulated trials concluded to non toxicity when the toxicity rate was truly 0.30.

Given these results, we wanted to explore more conservative rules for the designs combining efficacy assessment and toxicity monitoring; we explored $\pi_{tox} = 0.20, 0.225, 0.25$ and $\tau = 0.90, 0.95$ when associated with the TOP-design. Results are reported in the following paragraph.

Considering both designs for efficacy and toxicity

To assess whether the parameters $\pi_{tox} = 0.25$ and $\tau = 0.95$ in the PP_{tox} stopping rule ($Pr(p_{tox} > \pi_{tox} | D_n) > \tau$) lead to desirable properties, we evaluated 3 values of π_{tox} (0.20, 0.225 and 0.25) and 2 values of τ (0.90 and 0.95). The PP_{tox} design was combined with a TOP-design for efficacy and the parameters of the TOP-design were identical as for the "TOP_{eff} + PP_{tox}" design.

We took the same 10 scenarios presented in table 1 of the article and with a positive correlation between efficacy and toxicity, as previously. Each scenario consisted of a replication of 10000 trials with the multinomial distribution.

Figures S2 and S3 in the following illustrate the results.

Both pairs, $(\pi_{\text{tox}} = 0.20$ and $\tau = 0.95)$ and $(\pi_{\text{tox}} = 0.225$ and $\tau = 0.90)$, gave results appropriate for our settings. The type I error rate is controlled under 5% for both (0.56% and 0.80% respectively), as well as for the other evaluated sets. Though, these type I error rates might seem very low, one must take into consideration that both efficacy and safety of the drug are concomitantly evaluated and that, depending on the clinical settings, it may seem crucial to minimize the probability that a toxic drug goes through the trial successfully. This aspect could be discussed with clinicians when planning a trial using this approach. Compared to other pairs of parameters, in the scenarios simulated, these 2 sets have lower proportions of false rejection than the other sets except for $(\pi_{\text{tox}} = 0.20$ and $\tau = 0.90)$. Despite the higher power of $(\pi_{\text{tox}} = 0.225$ and $\tau = 0.90)$ over $(\pi_{\text{tox}} = 0.20$ and $\tau = 0.95)$ (83.3% vs 82.7%), the choice of the pair $(\pi_{\text{tox}} = 0.20$ and $\tau = 0.95)$ appeared the most appropriate in our setting. Our choice was motivated by the lower proportion of false positive in the toxic (scenarios 8, 9 and 10) and intermediate toxic (scenarios 3 and 4) cases for the second set of parameters. Because the attention on this study is particularly on avoiding to accept a toxic treatment, we prefer the rule that is more conservative in the situations where the toxicity is too high.

Between the 2 sets of parameters, the set $(\pi_{\text{tox}} = 0.20$ and $\tau = 0.95)$ enrolls slightly more patients than the set $(\pi_{\text{tox}} = 0.225$ and $\tau = 0.90)$ in all the scenarios we studied.

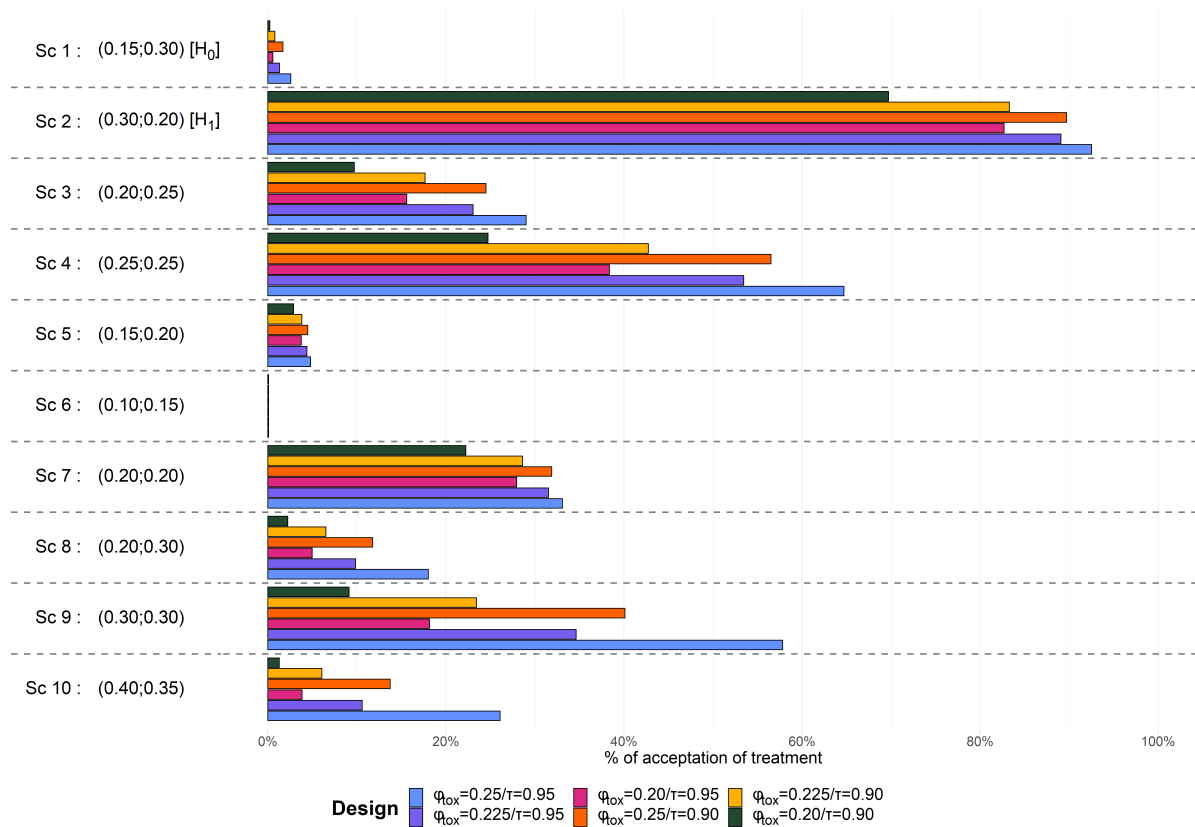


FIGURE S2 – Percentage of conclusion of efficacy and no toxicity in the 10 scenarios with a positive correlation between efficacy and toxicity for the different values of π_{tox} and τ . Numbers in parentheses represent p_{eff} and p_{tox} for each scenario.

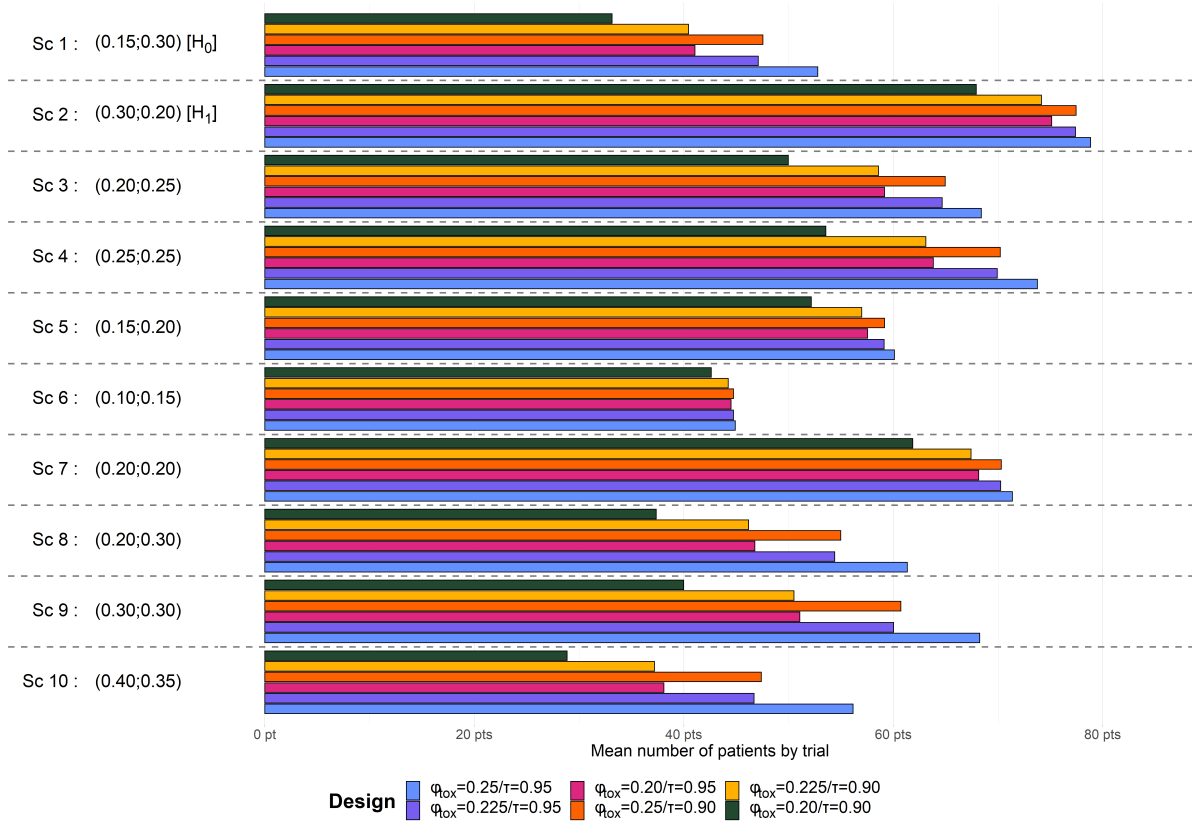


FIGURE S3 – Mean number of patients per trial in the 10 scenarios with a positive correlation between efficacy and toxicity for the different values of π_{tox} and τ . Numbers in parentheses represent p_{eff} and p_{tox} for each scenario.

Comparison with Simon’s two-stage design

For comparative purposes, we also compared the different approaches with a Simon’s 2 stage design combined with the PP_{tox} toxicity monitoring ($\pi_{tox} = 0.25$ and $\tau = 0.95$).

The settings used to optimize the Simon’s design were the following : $\alpha = 0.05$, $\beta = 0.10$, $p_u = 0.15$, $p_a = 0.30$. The optimal design found enrolled 82 with an interim analysis at 30 patients. If less than 6 responses occurred at the first stage, the trial was stopped for futility, as well as if less than 18 patients experienced a response during the whole trial.

Because the other designs we evaluated had analyses at 30 and 81 patients for efficacy, we forced identical sample sizes for the analyses in Simon’s design. Following the article (8), we determined optimal decision rules for this sample size, that is the design which minimizes the mean number of patients recruited in a trial when the treatment is not effective (with probability of response equal to p_u). Thus, the trial was stopped for futility at the 1st stage

if less than 5 responses occurred, and the treatment was considered unpromising if less than 18 responses occurred overall. This design (compared with the optimal design found without restriction on the number of enrolled patients) had a type I error rate of 4.87% (vs 4.61%), a power of 93.4% (vs 90.1%), an expected mean sample size under $p_u=0.15$ of 54.3 patients (vs 45.1) and a probability of early termination under p_u of 52.4% (vs 71.1%).

Design calibration

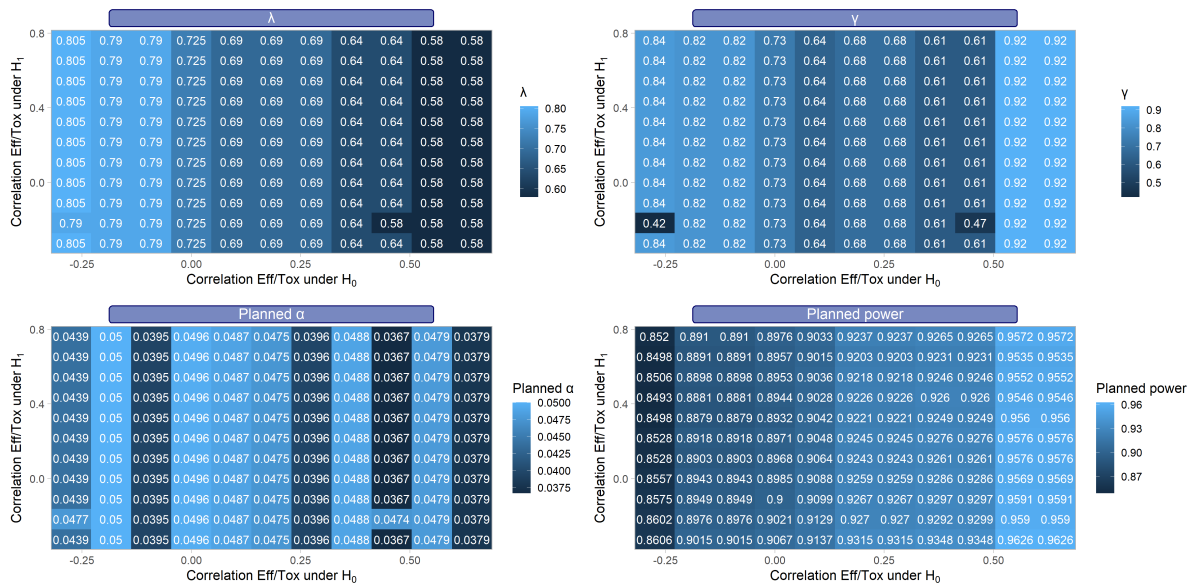


FIGURE S4 – Heatmap of the thresholds and a priori error rate relative to the correlation between efficacy and toxicity under H_0 and H_1 . The TOP-design was used with 2 interim analyses here.

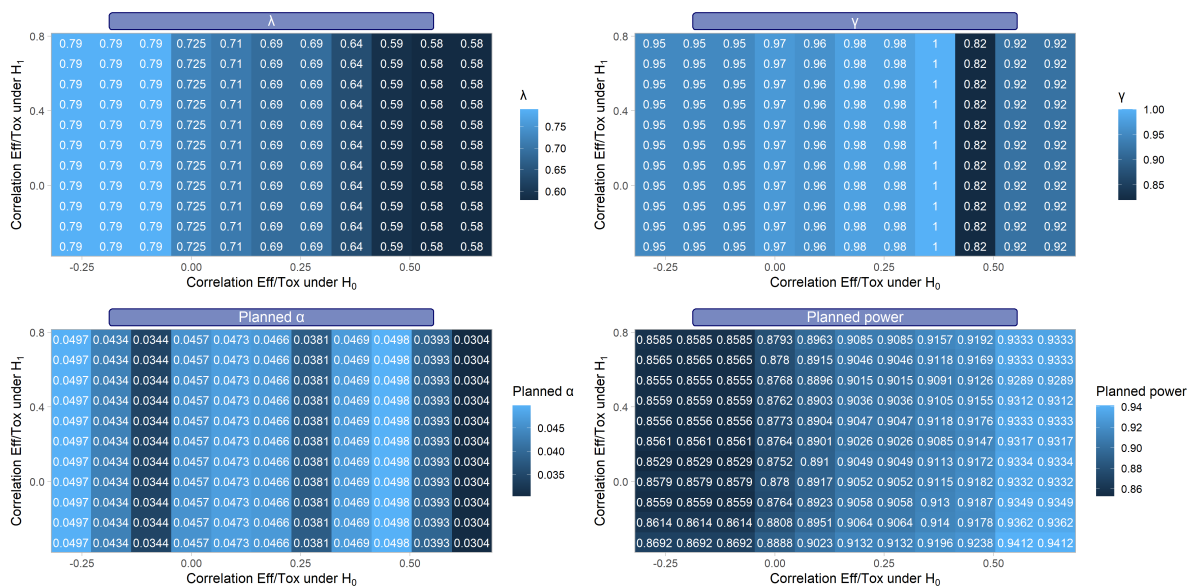


FIGURE S₅ – Heatmap of the thresholds and a priori error rate relative to the correlation between efficacy and toxicity under H_0 and H_1 . The TOP-design was used with 2 interim analyses for efficacy and 10 for toxicity as defined above.

3.2.2 Additional file 2 – Stopping rules for each design

TOP_{eff} + PP_{tox}

Decision rules for efficacy are the following :

TABLE S₁ – Stopping rules for futility of TOP_{eff} + PP_{tox} design.

Number of patients	Number of responses	Number of pending patients	Decision
30	≤ 3	≥ 12	Suspend inclusions
30	≤ 2	≤ 11	Stop for futility
30	3	≤ 11	Continue if ESS < 22.7 else stop for futility
30	≥ 4	≤ 26	Continue
81	≤ 81	≥ 1	Suspend inclusions
81	≤ 17	o	Conclude to futility
81	≥ 18	o	Conclude to efficacy

Decision rules for toxicity are the following (with the old rules in parentheses) :

TABLE S2 – Stopping rules for toxicity of TOP_{eff} + PP_{tox} design.

Number of patients	Stop for toxicity if > ... toxicities
5	2(2)
10	4(4)
15	5(6)
20	6(8)
30	9(11)
40	12(14)
50	14(17)
60	16(20)
70	19(23)
80	21(26)
90	24(29)

Simon + PP_{tox}

Decision rules for efficacy are the following :

TABLE S3 – Stopping rules for futility of Simon + PP_{tox} design.

Number of patients	Stop for futility if ≤ ... responses
30	4
81	17

Toxicity rules are the same as the TOP_{eff} + PP_{tox} design and are thus show in table S2.

BOP_{eff} + PP_{tox}

Efficacy rules are the following :

TABLE S4 – Stopping rules for futility of BOP_{eff} + PP_{tox} design.

Number of patients	Stop for futility if ≤ ... responses
30	3
81	17

Toxicity rules are the same as the TOP_{eff} + PP_{tox} design and are thus show in table S2.

TOP_{eff/tox} with 2 analyses

Efficacy rules are the following :

TABLE S5 – Stopping rules for futility of TOP_{eff/tox} design with 2 analyses.

Number of patients	Number of responses	Number of pending patients	Decision
30	≤ 3	≥ 12	Suspend inclusions
30	≤ 2	≤ 11	Stop for futility
30	3	≤ 11	Continue if ESS < 22.69 else stop for futility
30	≥ 4	≤ 26	Continue
81	≤ 81	≥ 1	Suspend inclusions
81	≤ 14	0	Conclude to futility
81	≥ 15	0	Conclude to efficacy

Toxicity rules are the following :

TABLE S6 – Stopping rules for toxicity of TOP_{eff/tox} design with 2 analyses.

Number of patients	Number of toxicities	Number of pending patients	Decision
30	≤ 10	≥ 12	Suspend inclusions
30	≤ 6	≤ 11	Continue
30	7	≤ 11	Continue if ESS > 20.23 else stop for toxicity
30	8	≤ 11	Continue if ESS > 23.37 else stop for toxicity
30	9	≤ 11	Continue if ESS > 26.52 else stop for toxicity
30	10	≤ 11	Continue if ESS > 29.67 else stop for toxicity
30	≥ 11	≤ 19	Stop for toxicity
81	≤ 81	≥ 1	Suspend inclusions
81	≤ 22	0	Conclude to acceptable toxicity
81	≥ 23	0	Conclude to toxicity

TOP_{eff/tox} with close monitoring of toxicity

Futility rules are the following :

TABLE S7 – Stopping rules for futility of TOP_{eff/tox} design with close monitoring of toxicity.

Number of patients	Number of responses	Number of pending patients	Decision
30	≤ 3	≥ 12	Suspend inclusions
30	≤ 2	≤ 11	Stop for futility
30	3	≤ 11	Continue if ESS < 25.9 else stop for futility
30	≥ 4	≤ 26	Continue
81	≤ 81	≥ 1	Suspend inclusions
81	≤ 14	0	Conclude to futility
81	≥ 15	0	Conclude to efficacy

Toxicity rules are the following :

TABLE S8 – Stopping rules for toxicity of TOP_{eff/tox} design with close monitoring of toxicity (1/2).

Patients	Toxicities	Pending patients	Toxic
5	≤ 3	≥ 1	Suspend inclusions
5	≤ 3	≤ 0	Continue
5	≥ 4	≤ 1	Stop for toxicity
10	≤ 5	≥ 2	Suspend inclusions
10	≤ 4	≤ 1	Continue
10	5	≤ 1	Stop for toxicity if ESS ≤ 9.52
10	≥ 6	≤ 4	Stop for toxicity
15	≤ 6	≥ 3	Suspend inclusions
15	≤ 5	≤ 2	Continue
15	6	≤ 2	Stop for toxicity if ESS ≤ 13.12
15	≥ 7	≤ 8	Stop for toxicity
20	≤ 8	≥ 5	Suspend inclusions
20	≤ 6	≤ 4	Continue
20	7	≤ 4	Stop for toxicity if ESS ≤ 16.85
20	8	≤ 4	Stop for toxicity if ESS ≤ 19.7
20	≥ 9	≤ 11	Stop for toxicity
30	≤ 10	≥ 12	Suspend inclusions
30	≤ 6	≤ 11	Continue
30	7	≤ 11	Stop for toxicity if ESS ≤ 18.57
30	8	≤ 11	Stop for toxicity if ESS ≤ 21.58
30	9	≤ 11	Stop for toxicity if ESS ≤ 24.6
30	10	≤ 11	Stop for toxicity if ESS ≤ 27.65
30	≥ 11	≤ 19	Stop for toxicity
40	≤ 13	≥ 20	Suspend inclusions
40	≤ 6	≤ 19	Continue
40	7	≤ 19	Stop for toxicity if ESS ≤ 20.13
40	8	≤ 19	Stop for toxicity if ESS ≤ 23.26
40	9	≤ 19	Stop for toxicity if ESS ≤ 26.4
40	10	≤ 19	Stop for toxicity if ESS ≤ 29.55
40	11	≤ 19	Stop for toxicity if ESS ≤ 32.71
40	12	≤ 19	Stop for toxicity if ESS ≤ 35.88
40	13	≤ 19	Stop for toxicity if ESS ≤ 39.06
40	≥ 14	≤ 26	Stop for toxicity
50	≤ 15	≥ 31	Suspend inclusions
50	≤ 6	≤ 30	Continue
50	7	≤ 30	Stop for toxicity if ESS ≤ 21.63
50	8	≤ 30	Stop for toxicity if ESS ≤ 24.87
50	9	≤ 30	Stop for toxicity if ESS ≤ 28.12
50	10	≤ 30	Stop for toxicity if ESS ≤ 31.38
50	11	≤ 30	Stop for toxicity if ESS ≤ 34.63
50	12	≤ 30	Stop for toxicity if ESS ≤ 37.89
50	13	≤ 30	Stop for toxicity if ESS ≤ 41.16
50	14	≤ 30	Stop for toxicity if ESS ≤ 44.42
50	15	≤ 30	Stop for toxicity if ESS ≤ 47.69
50	≥ 16	≤ 34	Stop for toxicity

TABLE S9 – Stopping rules for toxicity of TOP_{eff/tox} design with close monitoring of toxicity (2/2).

Patients	Toxicities	Pending patients	Toxic
60	≤ 18	≥ 45	Suspend inclusions
60	≤ 4	≤ 44	Continue
60	5	≤ 44	Stop for toxicity if ESS ≤ 16.45
60	6	≤ 44	Stop for toxicity if ESS ≤ 19.8
60	7	≤ 44	Stop for toxicity if ESS ≤ 23.16
60	8	≤ 44	Stop for toxicity if ESS ≤ 26.51
60	9	≤ 44	Stop for toxicity if ESS ≤ 29.86
60	10	≤ 44	Stop for toxicity if ESS ≤ 33.21
60	11	≤ 44	Stop for toxicity if ESS ≤ 36.55
60	12	≤ 44	Stop for toxicity if ESS ≤ 39.9
60	13	≤ 44	Stop for toxicity if ESS ≤ 43.25
60	14	≤ 44	Stop for toxicity if ESS ≤ 46.6
60	15	≤ 44	Stop for toxicity if ESS ≤ 49.94
60	16	≤ 44	Stop for toxicity if ESS ≤ 53.29
60	17	≤ 44	Stop for toxicity if ESS ≤ 56.63
60	18	≤ 44	Stop for toxicity if ESS ≤ 59.98
60	≥ 19	≤ 41	Stop for toxicity
70	≤ 20	≥ 61	Suspend inclusions
70	≤ 2	≤ 60	Continue
70	3	≤ 60	Stop for toxicity if ESS ≤ 10.81
70	4	≤ 60	Stop for toxicity if ESS ≤ 14.33
70	5	≤ 60	Stop for toxicity if ESS ≤ 17.82
70	6	≤ 60	Stop for toxicity if ESS ≤ 21.3
70	7	≤ 60	Stop for toxicity if ESS ≤ 24.77
70	8	≤ 60	Stop for toxicity if ESS ≤ 28.23
70	9	≤ 60	Stop for toxicity if ESS ≤ 31.69
70	10	≤ 60	Stop for toxicity if ESS ≤ 35.13
70	11	≤ 60	Stop for toxicity if ESS ≤ 38.57
70	12	≤ 60	Stop for toxicity if ESS ≤ 42.01
70	13	≤ 60	Stop for toxicity if ESS ≤ 45.44
70	14	≤ 60	Stop for toxicity if ESS ≤ 48.87
70	15	≤ 60	Stop for toxicity if ESS ≤ 52.29
70	16	≤ 60	Stop for toxicity if ESS ≤ 55.71
70	17	≤ 60	Stop for toxicity if ESS ≤ 59.13
70	18	≤ 60	Stop for toxicity if ESS ≤ 62.55
70	19	≤ 60	Stop for toxicity if ESS ≤ 65.96
70	20	≤ 60	Stop for toxicity if ESS ≤ 69.37
70	≥ 21	≤ 49	Stop for toxicity
81	≤ 81	≥ 1	Suspend inclusions
81	≥ 23	0	Stop for toxicity
81	≤ 22	0	Continue

3.2.3 Additional file 3 – Additional results

Operating characteristics

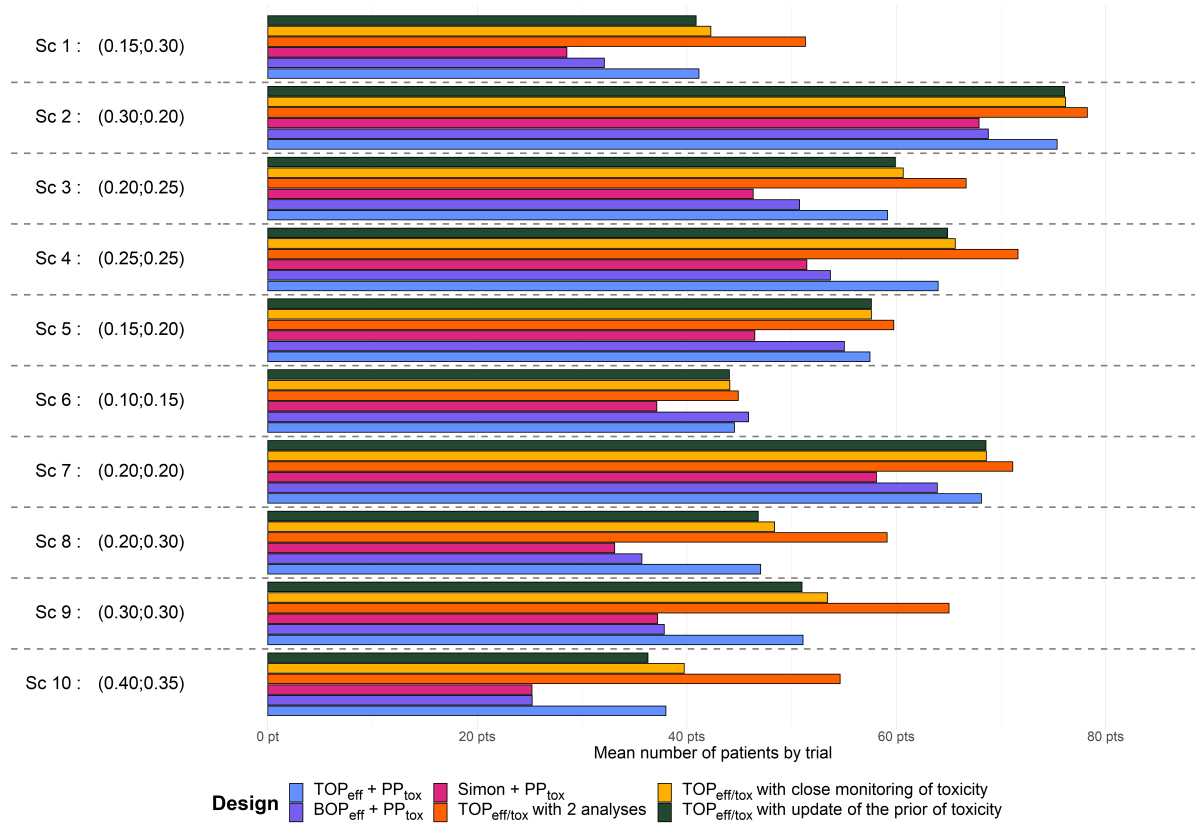


FIGURE S6 – Average number of patients per trial in the 10 scenarios with a positive correlation between efficacy and toxicity. Numbers in parentheses represent p_{eff} and p_{tox} for each scenario.

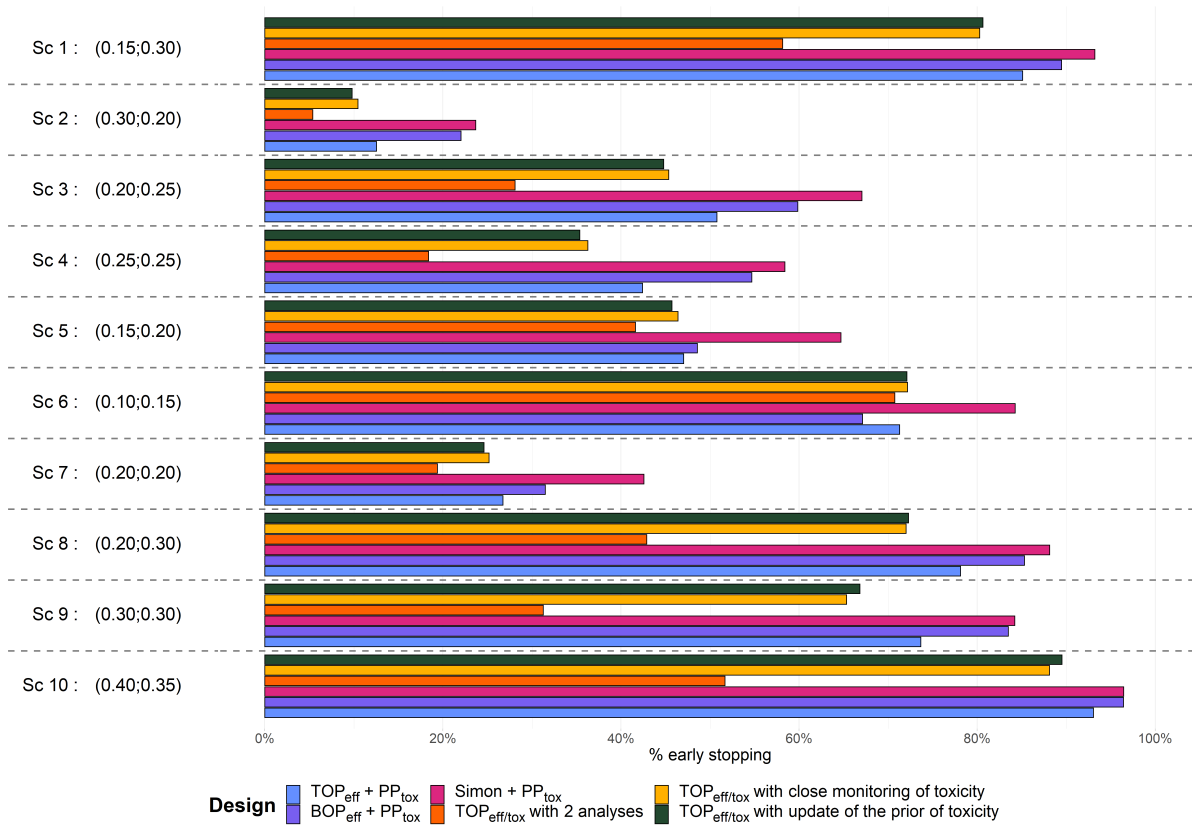


FIGURE S7 – Proportion of early stopping in the 10 scenarios with a positive correlation between efficacy and toxicity. Numbers in parentheses represent p_{eff} and p_{tox} for each scenario.

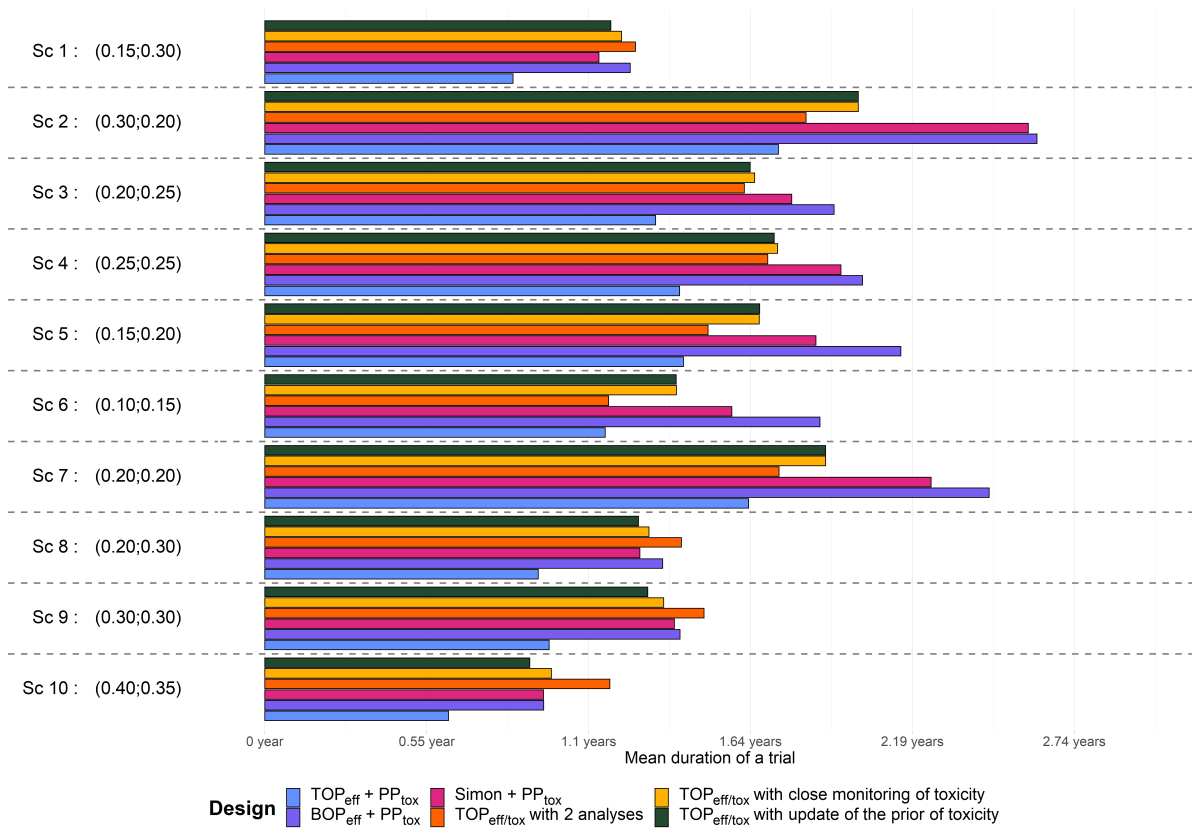


FIGURE S8 – Mean duration of a trial in the 10 scenarios with a positive correlation between efficacy and toxicity. Numbers in parentheses represent p_{eff} and p_{tox} for each scenario.

Correlation between efficacy and toxicity

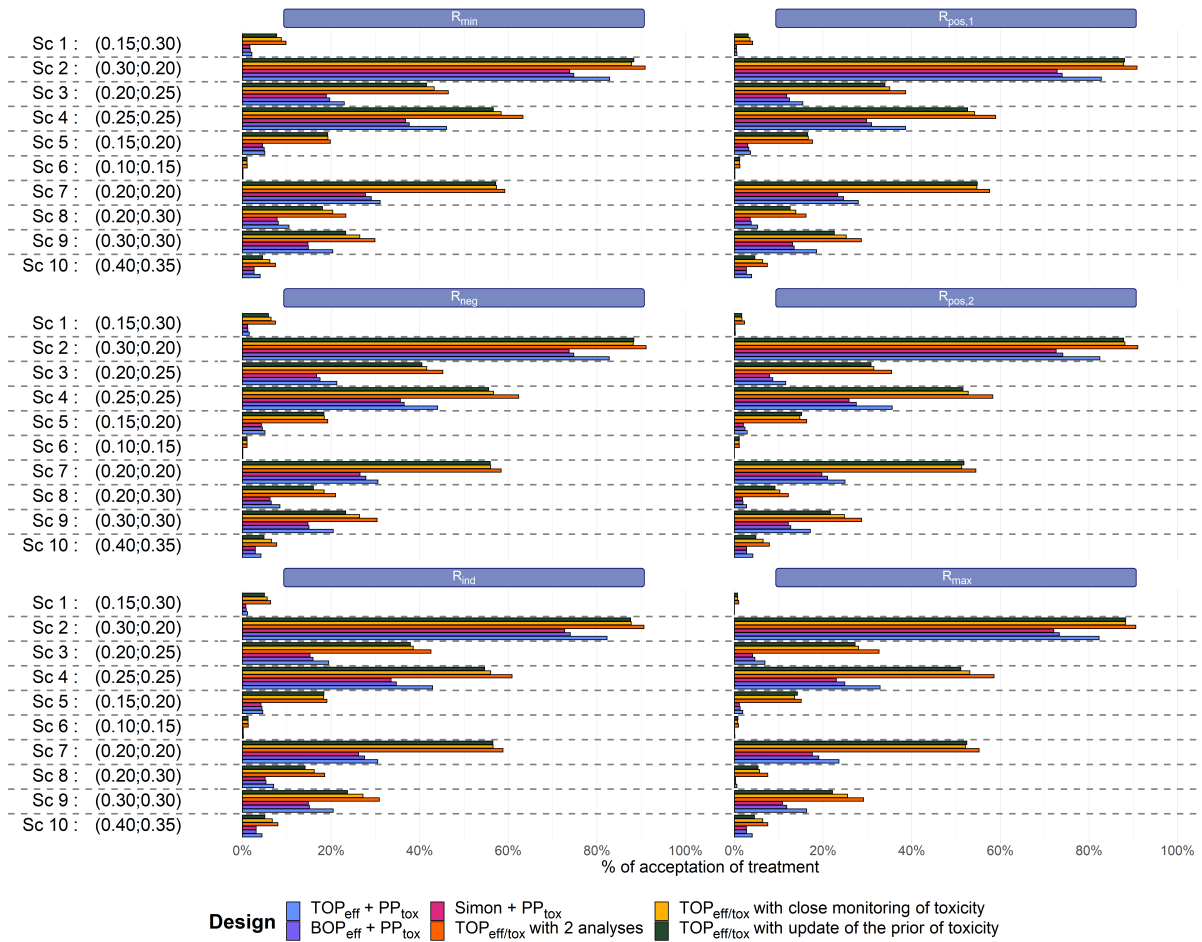


FIGURE S9 – Percentage of conclusion of efficacy and no toxicity in the 10 scenarios with 6 sets of correlation between efficacy and toxicity. Numbers in parentheses represent p_{eff} and p_{tox} for each scenario.

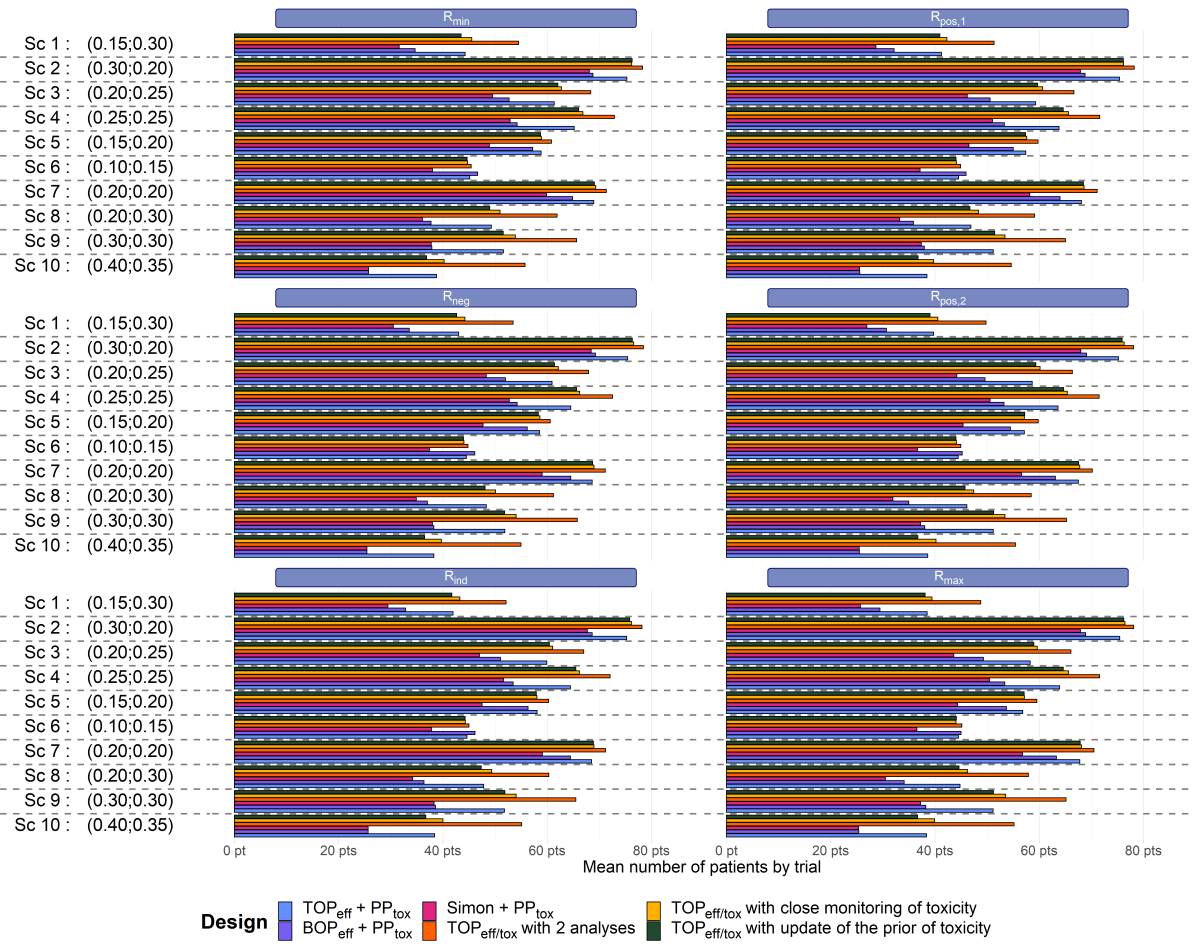


FIGURE S10 – Mean number of patients per trial in the 10 scenarios with 6 sets of correlation between efficacy and toxicity. Numbers in parentheses represent p_{eff} and p_{tox} for each scenario.

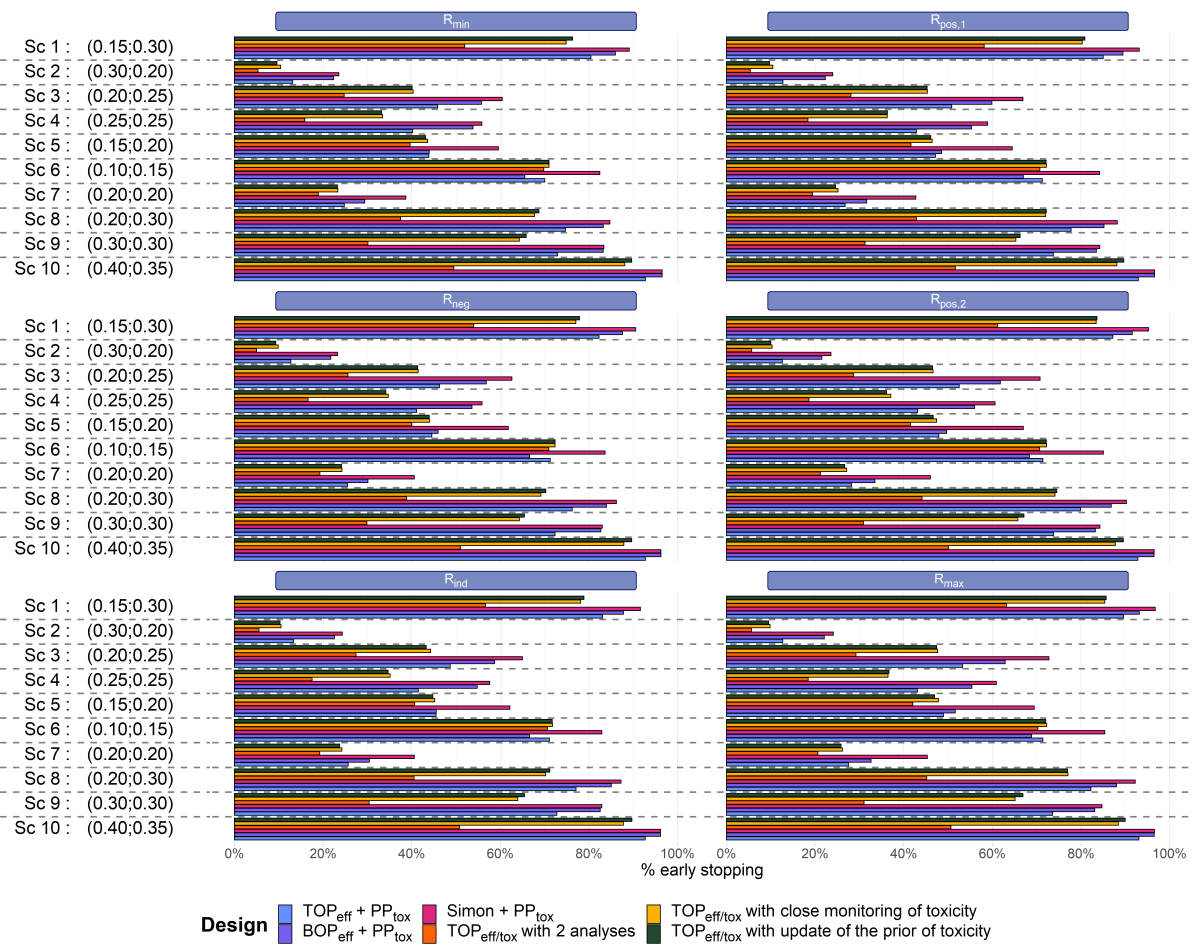


FIGURE S11 – Proportion of early stopping in the 10 scenarios with 6 sets of correlation between efficacy and toxicity. Numbers in parentheses represent p_{eff} and p_{tox} for each scenario.

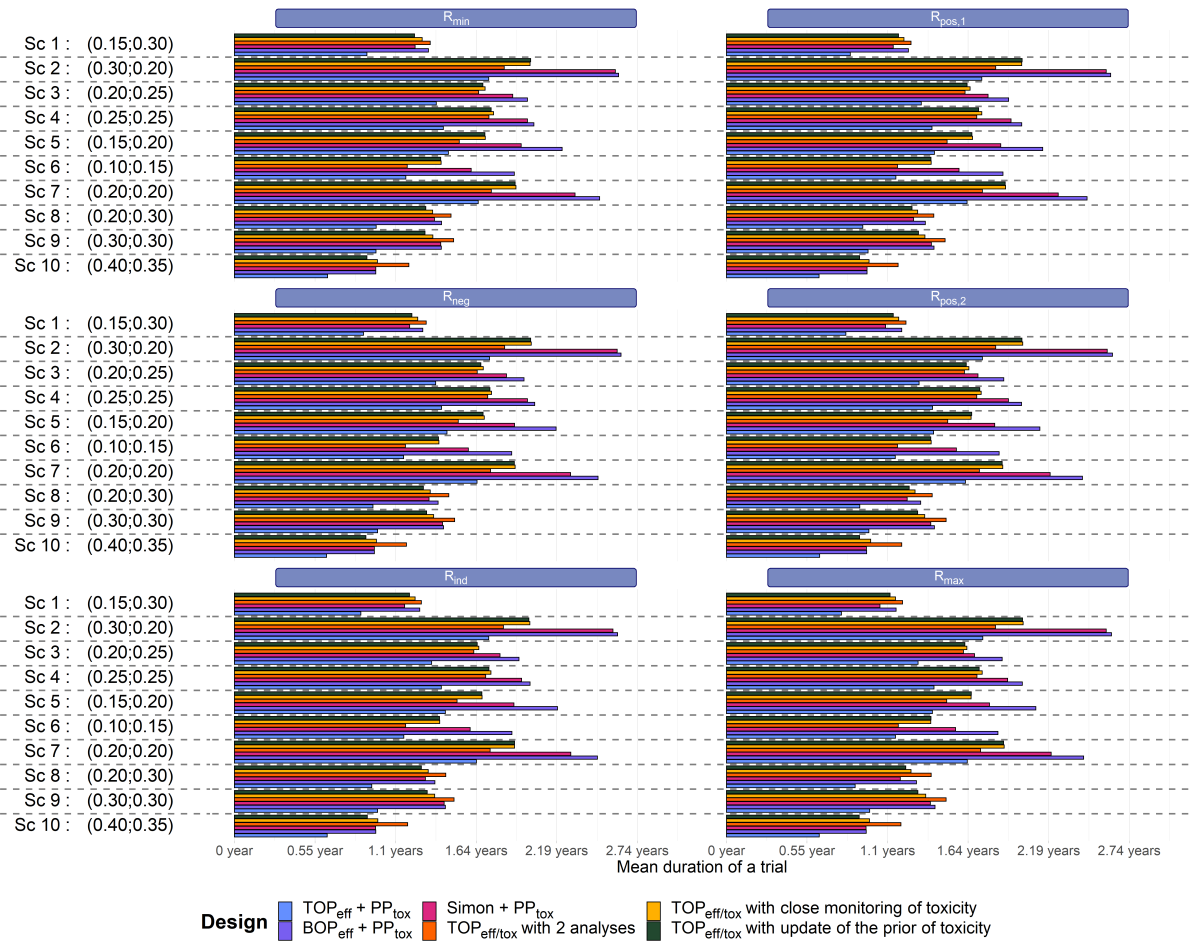


FIGURE S12 – Mean duration of a trial in the 10 scenarios with 6 sets of correlation between efficacy and toxicity. Numbers in parentheses represent p_{eff} and p_{tox} for each scenario.

4 Conclusion

Nous avons comparé différents schémas adaptatifs séquentiels pour des essais cliniques de phase II, adaptés aux caractéristiques des nouveaux traitements en oncologie, telles que les immunothérapies et les thérapies ciblées : fenêtre d'activité retardée, risque de toxicité non négligeable (62). En particulier, nous avons considéré des schémas adaptés pour un essai de phase II actuellement en période d'inclusion, évaluant la combinaison d'un inhibiteur de tyrosine kinase (petite molécule) et d'un inhibiteur de point de contrôle immunitaire (immunothérapie) dans le cancer oesogastrique avancé, l'essai METIMGAST (NCT05135845). Des signaux sur le profil de tolérance ayant été observés dans un autre essai évaluant l'un des deux traitements dans une autre indication, la prise en compte de la toxicité est apparue nécessaire pour cet essai de phase II. Les schémas comparés devaient donc permettre d'une part une analyse intermédiaire sur le critère d'efficacité avec une règle d'arrêt pour futilité, et d'autre part un suivi rapproché de la tolérance, avec règles d'arrêt précoce pour toxicité.

Nous avons considéré des schémas de phase II combinant une évaluation de l'efficacité avec soit le schéma de Simon (8) soit une approche bayésienne (47; 49), et une évaluation de la toxicité avec règle d'arrêt précoce avec une approche bayésienne (11; 47; 48). Les schémas bayésiens ont l'avantage de proposer une approche globale et flexible pour l'évaluation conjointe de l'efficacité et la toxicité en s'appuyant sur un modèle multinomial (63). De plus, le schéma TOP permet de prendre en compte les observations encore incomplètes lors des analyses intermédiaires donc d'éviter l'interruption des inclusions, contrairement au schéma de Simon (47). La durée moyenne des essais réalisés avec ce schéma en est ainsi réduite.

Notre étude de simulations montre que le schéma TOP utilisé pour l'évaluation de l'efficacité avec un monitoring rapproché de la toxicité distinct assure le contrôle du risque de faux positifs (conclusion à tort à un traitement efficace et non toxique) et présente une bonne puissance (conclusion à raison à un traitement efficace et non toxique). La planification de l'essai doit cependant prendre en compte qu'à effectif maximal constant, l'augmentation du nombre d'analyses intermédiaires de toxicité conduit à diminuer la puissance globale de l'essai.

De plus, dans un contexte d'évaluation conjointe de l'efficacité et la toxicité, le schéma

TOP avec critères multiples (efficacité + toxicité) permet de prendre en compte directement la corrélation entre ces 2 critères, via l'utilisation d'un modèle multinomial. Une corrélation positive est en règle générale observée en oncologie (64; 65). Toutefois, la surestimation de cette corrélation lors de la planification de l'essai entraîne un risque d'augmentation du risque de faux positifs.

Enfin, nous avons adapté le schéma TOP avec critères multiples (efficacité + toxicité), afin de prendre en compte toute l'information disponible concernant la toxicité dans le cadre de l'essai METIMGAST (cohorte c-MET+) via une distribution *a priori* informative pour la probabilité de toxicité dans le modèle. D'une manière générale, l'approche bayésienne apporte la possibilité d'enrichir l'analyse d'un essai avec de l'information extérieure c'est-à-dire de prendre une décision d'évaluation en tenant compte de l'ensemble des données et connaissances disponibles à un moment donné (66).

En conclusion, les spécificités des nouvelles thérapies en oncologie peuvent nécessiter l'utilisation de schémas d'essais cliniques innovants pour leur évaluation optimale, en termes de risque d'erreur et de durée d'essai notamment. Les approches bayésiennes semblent particulièrement adaptées à ce contexte : elles correspondent à une démarche intuitive intégrant les connaissances disponibles *a priori* pour l'évaluation de traitements (66; 67). Sous réserve d'un travail de calibration préalable des schémas pour assurer leurs propriétés statistiques et performances, il s'agit de méthodes flexibles permettant la planification de schémas d'essais adaptés spécifiquement à un contexte clinique.

Le Doyen

Pr Djillali ANNANE

Le Président du jury

Pr Sylvie CHEVRET

Le Président de l'Université

Pr Alain BUI



Bibliographie

- [1] Inserm. Développement du médicament : de l'éprouvette à la pharmacie ; 2017. Available from : <https://www.inserm.fr/dossier/medicament-developpement/>.
- [2] Bryant J, Day R. Incorporating toxicity considerations into the design of two-stage phase II clinical trials. *Biometrics*. 1995 ;1372-1383.
- [3] Le Tourneau C, Lee JJ, Siu LL. Dose escalation methods in phase I cancer clinical trials. *JNCI : Journal of the National Cancer Institute*. 2009;101(10) :708-720.
- [4] O'Quigley J, Pepe M, Fisher L. Continual reassessment method : a practical design for phase I clinical trials in cancer. *Biometrics*. 1990 :33-48.
- [5] Ji Y, Wang SJ. Modified toxicity probability interval design : a safer and more reliable method than the $3+3$ design for practical phase I trials. *Journal of Clinical Oncology*. 2013;31(14) :1785.
- [6] Mander AP, Sweeting MJ. A product of independent beta probabilities dose escalation design for dual-agent phase I trials. *Statistics in medicine*. 2015;34(8) :1261-1276.
- [7] Yuan Y, Hess KR, Hilsenbeck SG, Gilbert MR. Bayesian Optimal Interval Design : A Simple and Well-Performing Design for Phase I Oncology Trials BOIN : A Simple, Well-Performing Bayesian Phase I Design. *Clinical Cancer Research*. 2016;22(17) :4291-4301.
- [8] Simon R. Optimal two-stage designs for phase II clinical trials. *Controlled clinical trials*. 1989;10(1) :1-10.
- [9] Sambucini V. Efficacy and toxicity monitoring via Bayesian predictive probabilities in phase II clinical trials. *Statistical Methods & Applications*. 2021;30(2) :637-663.
- [10] Conaway MR, Petroni GR. Designs for phase II trials allowing for a trade-off between response and toxicity. *Biometrics*. 1996 :1375-1386.
- [11] Thall PF, Simon RM, Estey EH. New statistical strategy for monitoring safety and efficacy in single-arm clinical trials. *Journal of Clinical Oncology*. 1996;14(1) :296-303.
- [12] Hobbs BP, Chen N, Lee JJ. Controlled multi-arm platform design using predictive probability. *Statistical methods in medical research*. 2018;27(1) :65-78.
- [13] Estey EH, Thall PF. New designs for phase 2 clinical trials. *Blood*. 2003;102(2) :442-448.
- [14] Wages NA, Tait C. Seamless phase I/II adaptive design for oncology trials of molecularly targeted agents. *Journal of biopharmaceutical statistics*. 2015;25(5) :903-920.
- [15] Hoering A, LeBlanc M, Crowley J. Seamless phase I-II trial design for assessing toxicity and efficacy for targeted agents. *Clinical Cancer Research*. 2011;17(4) :640-646.
- [16] Mozgunov P, Jaki T. An information theoretic phase I-II design for molecularly targeted agents that does not require an assumption of monotonicity. *Journal of the Royal Statistical Society Series C, Applied Statistics*. 2019;68(2) :347.

- [17] Nebiyou Bekele B, Shen Y. A Bayesian approach to jointly modeling toxicity and biomarker expression in a phase I/II dose-finding trial. *Biometrics*. 2005;61(2):343–354.
- [18] Bretz F, Schmidli H, König F, Racine A, Maurer W. Confirmatory seamless phase II/III clinical trials with hypotheses selection at interim : general concepts. *Biometrical Journal : Journal of Mathematical Methods in Biosciences*. 2006;48(4):623–634.
- [19] Wang M, Dignam JJ, Zhang QE, DeGroot JF, Mehta MP, Hunsberger S. Integrated phase II/III clinical trials in oncology : A case study. *Clinical Trials*. 2012;9(6):741–747.
- [20] Chapple AG, Thall PF. A hybrid phase I-II/III clinical trial design allowing dose re-optimization in phase III. *Biometrics*. 2019;75(2):371–381.
- [21] Li Q, Lin J, Lin Y. Adaptive design implementation in confirmatory trials : methods, practical considerations and case studies. *Contemporary Clinical Trials*. 2020;98:106096.
- [22] for Medicinal Products for Human Use C, et al. Reflection paper on methodological issues in confirmatory clinical trials planned with an adaptive design. London : EMEA. 2007.
- [23] Collignon O, Koenig F, Koch A, Hemmings RJ, Pétavy F, Saint-Raymond A, et al. Adaptive designs in clinical trials : from scientific advice to marketing authorisation to the European Medicine Agency. *Trials*. 2018;19(1):1–14.
- [24] FDA. Adaptive Designs for Clinical Trials of Drugs and Biologics. Center for Biologics Evaluation and Research (CBER). 2019.
- [25] FDA. Development Approval Process | Drugs; 2019. Available from : <https://www.fda.gov/drugs/development-approval-process-drugsDeveloping>.
- [26] Ghisoni E, Wicky A, Bouchaab H, Imbimbo M, Delyon J, Moura BG, et al. Late-onset and long-lasting immune-related adverse events from immune checkpoint-inhibitors : An overlooked aspect in immunotherapy. *European journal of cancer*. 2021;149:153–164.
- [27] Garsen M, Steenhof M, Zwiers A. A decade of marketing authorization applications of anticancer drugs in the European Union : an analysis of procedural timelines. *Therapeutic Innovation & Regulatory Science*. 2021;55(4):633–642.
- [28] Sharma P, Wagner K, Wolchok J, Allison J. Novel cancer immunotherapy agents with survival benefit : recent successes and next steps. *Nat Rev Cancer*. 2011;11(11):805–812.
- [29] Postel-Vinay S, Aspeslagh S, Lanoy E, Robert C, Soria JC, Marabelle A. Challenges of phase I clinical trials evaluating immune checkpoint-targeted antibodies. *Annals of Oncology*. 2016;27(2):214–224.
- [30] Fraisse J, Dinart D, Tosi D, Bellera C, Mollevi C. Optimal biological dose : a systematic review in cancer phase I clinical trials. *BMC cancer*. 2021;21(1):1–10.

- [31] Li Q, Lin J, Liu M, Wu L, Liu Y. Utilizing Surrogate Endpoints in Adaptive Designs with Delayed Treatment Effect. *Statistics in Biopharmaceutical Research*. 2021;(just-accepted) :1–15.
- [32] Gérard L, Manfredi S. Epidemiology and risk factors for stomach and oesogastric junction cancer. *Hépatogastro & Oncologie Digestive*. 2019;26(6) :565–572.
- [33] Bray F, Ferlay J, Soerjomataram I, Siegel R, Torre L, Jemal A. Global cancer statistics 2018 : GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries (vol 68, pg 394, 2018). *Ca-a Cancer Journal for Clinicians*. 2020;70(4) :313–313.
- [34] Lemoine N, Adenis A, Bouche O, Duhamel A, Heurgue A, Leteurtre E, et al. Signet ring cells and efficacy of first-line chemotherapy in advanced gastric or oesogastric junction adenocarcinoma. *Anticancer research*. 2016;36(10) :5543–5549.
- [35] Al-Batran SE, Homann N, Pauligk C, Goetze TO, Meiler J, Kasper S, et al. Perioperative chemotherapy with fluorouracil plus leucovorin, oxaliplatin, and docetaxel versus fluorouracil or capecitabine plus cisplatin and epirubicin for locally advanced, resectable gastric or gastro-oesophageal junction adenocarcinoma (FLOT4) : a randomised, phase 2/3 trial. *The Lancet*. 2019;393(10184) :1948–1957.
- [36] Thuss-Patience PC, Kretzschmar A, Bichev D, Deist T, Hinke A, Breithaupt K, et al. Survival advantage for irinotecan versus best supportive care as second-line chemotherapy in gastric cancer—a randomised phase III study of the Arbeitsgemeinschaft Internistische Onkologie (AIO). *European journal of cancer*. 2011;47(15) :2306–2314.
- [37] Kang JH, Lee SI, Lim D, Park K, Oh S, Kwon H, et al. Salvage chemotherapy for pretreated gastric cancer : a randomized phase III trial comparing chemotherapy plus best supportive care with best supportive care alone. *Clin Oncol*. 2012;30(13) :1513–1518.
- [38] Ford HE, Marshall A, Bridgewater JA, Janowitz T, Coxon FY, Wadsley J, et al. Docetaxel versus active symptom control for refractory oesophagogastric adenocarcinoma (COUGAR-02) : an open-label, phase 3 randomised controlled trial. *The lancet oncology*. 2014;15(1) :78–86.
- [39] Shitara K, Van Cutsem E, Bang YJ, Fuchs C, Wyrwicz L, Lee KW, et al. Efficacy and safety of pembrolizumab or pembrolizumab plus chemotherapy vs chemotherapy alone for patients with first-line, advanced gastric cancer : the KEYNOTE-062 phase 3 randomized clinical trial. *JAMA oncology*. 2020;6(10) :1571–1580.
- [40] Böger C, Behrens HM, Mathiak M, Krüger S, Kalthoff H, Röcken C. PD-L1 is an independent prognostic predictor in gastric cancer of Western patients. *Oncotarget*. 2016;7(17) :24269.
- [41] Li H, Li CW, Li X, Ding Q, Guo L, Liu S, et al. MET inhibitors promote liver tumor evasion of the immune response by stabilizing PDL1. *Gastroenterology*. 2019;156(6) :1849–1861.

- [42] Glodde N, Bald T, van den Boorn-Konijnenberg D, Nakamura K, O'Donnell JS, Szczepanski S, et al. Reactive neutrophil responses dependent on the receptor tyrosine kinase c-MET limit cancer immunotherapy. *Immunity*. 2017;47(4) :789-802.
- [43] Lee J, Tran P, Klempner S. Targeting the MET pathway in gastric and oesophageal cancers : refining the optimal approach. *Clinical oncology*. 2016;28(8) :e35-e44.
- [44] Van Cutsem E, Karaszewska B, Kang YK, Chung HC, Shankaran V, Siena S, et al. A multicenter phase II study of AMG 337 in patients with MET-amplified gastric/gastroesophageal junction/esophageal adenocarcinoma and other MET-amplified solid tumors. *Clinical Cancer Research*. 2019;25(8) :2414-2423.
- [45] Aparicio T, Cozic N, de la Fouchardière C, Meriaux E, Plaza J, Mineur L, et al. The activity of crizotinib in chemo-refractory MET-amplified esophageal and gastric adenocarcinomas : results from the acsé-crizotinib program. *Targeted Oncology*. 2021;16(3) :381-388.
- [46] Morris TP, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods. *Statistics in medicine*. 2019;38(11) :2074-2102.
- [47] Lin R, Coleman RL, Yuan Y. TOP : Time-to-event Bayesian optimal phase II trial design for cancer immunotherapy. *JNCI : Journal of the National Cancer Institute*. 2020;112(1) :38-45.
- [48] Ivanova A, Qaqish BF, Schell MJ. Continuous toxicity monitoring in phase II trials in oncology. *Biometrics*. 2005;61(2) :540-545.
- [49] Zhou H, Lee JJ, Yuan Y. BOP₂ : Bayesian optimal design for phase II clinical trials with simple and complex endpoints. *Statistics in medicine*. 2017;36(21) :3302-3314.
- [50] Geller NL, Follman D, Leifer D, Carter SL. Design of early trials in stem cell transplantation : a hybrid frequentist-Bayesian approach. *Advances in Clinical Trial Biostatistics*. 2003 :41.
- [51] Dufour JM, Neves J. Chapter 1 - Finite-sample inference and nonstandard asymptotics with Monte Carlo tests and R. In : Vinod HD, Rao CR, editors. *Conceptual Econometrics Using R*. vol. 41 of *Handbook of Statistics*. Elsevier; 2019. p. 3-31.
- [52] Spiegelhalter DJ, Abrams KR, Myles JP. *Bayesian approaches to clinical trials and health-care evaluation*. vol. 13. John Wiley & Sons; 2004.
- [53] Jiang L, Yan F, Thall PF, Huang X. Comparing Bayesian early stopping boundaries for phase II clinical trials. *Pharmaceutical statistics*. 2020;19(6) :928-939.
- [54] Puzanov I, Diab A, Abdallah K, Bingham Cr, Brogdon C, Dadu R, et al. Managing toxicities associated with immune checkpoint inhibitors : consensus recommendations from the Society for Immunotherapy of Cancer (SITC) Toxicity Management Working Group. *Journal for immunotherapy of cancer*. 2017;5(1) :1-28.

- [55] Braun TM, Levine JE, Ferrara JL. Determining a maximum tolerated cumulative dose : dose reassignment within the TITE-CRM. *Controlled clinical trials*. 2003;24(6) :669–681.
- [56] Cheung YK, Chappell R. Sequential designs for phase I clinical trials with late-onset toxicities. *Biometrics*. 2000;56(4) :1177–1182.
- [57] Morita S, Thall PF, Müller P. Determining the effective sample size of a parametric prior. *Biometrics*. 2008;64(2) :595–602.
- [58] Cameron D, Massie C, Kerr G, Leonard R. Moderate neutropenia with adjuvant CMF confers improved survival in early breast cancer. *British journal of cancer*. 2003;89(10) :1837–1842.
- [59] Saarto T, Blomqvist C, Rissanen P, Auvinen A, Elomaa I. Haematological toxicity : a marker of adjuvant chemotherapy efficacy in stage II and III breast cancer. *British journal of cancer*. 1997;75(2) :301–305.
- [60] Palmieri DJ, Carlino MS. Immune checkpoint inhibitor toxicity. *Current oncology reports*. 2018;20(9) :1–12.
- [61] Das S, Johnson DB. Immune-related adverse events and anti-tumor efficacy of immune checkpoint inhibitors. *Journal for immunotherapy of cancer*. 2019;7(1) :1–11.
- [62] de Miguel M, Calvo E. Early-phase clinical drug development of novel agents : a changing paradigm. *Annals of Oncology*. 2019;30(7) :1033–1037. Targeting the PI3-kinase pathway in triple-negative breast cancer.
- [63] Yuan Y, Lee JJ, Hilsenbeck SG. Model-assisted designs for early-phase clinical trials : simplicity meets superiority. *JCO Precision Oncology*. 2019;3 :1–12.
- [64] Hussaini S, Chehade R, Boldt RG, Raphael J, Blanchette P, Vareki SM, et al. Association between immune-related side effects and efficacy and benefit of immune checkpoint inhibitors—a systematic review and meta-analysis. *Cancer treatment reviews*. 2021;92 :102134.
- [65] Pinato DJ, Marron TU, Mishra-Kalyani PS, Gong Y, Wei G, Szafron D, et al. Treatment-related toxicity and improved outcome from immunotherapy in hepatocellular cancer : Evidence from an FDA pooled analysis of landmark clinical trials with validation from routine practice. *European Journal of Cancer*. 2021;157 :140–152.
- [66] Su L, Chen X, Zhang J, Yan F. Comparative Study of Bayesian Information Borrowing Methods in Oncology Clinical Trials. *JCO Precision Oncology*. 2022;(6) :e2100394. PMID : 35263169.
- [67] Lin R, Lee JJ. In : Bekker A, Chen DDG, Ferreira JT, editors. *Novel Bayesian Adaptive Designs and Their Applications in Cancer Clinical Trials*. Cham : Springer International Publishing; 2020. p. 395–426. Available from : https://doi.org/10.1007/978-3-030-42196-0_17.

TITRE : Comparaison de schémas d'essais de phase II pour l'évaluation des nouveaux traitements en cancérologie

Introduction : Les nouvelles thérapeutiques en cancérologie telles que les immunothérapies et les thérapies ciblées remettent en cause les schémas usuels d'essais cliniques aux différentes phases de développement du médicament, du fait de leurs mécanismes d'action différents des chimiothérapies cytotoxiques. Nous avons pris l'exemple de la planification d'un essai de phase II évaluant une combinaison de nouvelles thérapies pour le traitement de l'adénocarcinome oesogastrique avancé nécessitant la définition d'un schéma adaptatif : fenêtre d'observation prolongée et évaluation conjointe et séquentielle de l'efficacité et de la toxicité.

Méthodes : Nous avons défini 6 schémas adaptatifs candidats par combinaison des approches suivantes : schéma bayésien permettant ou non un recrutement continu de patients, *monitoring* bayésien de la probabilité de toxicité séparé de l'évaluation d'efficacité ou évaluation conjointe des 2 critères selon un modèle multinomial, schéma de Simon en 2 étapes, intégration de données de toxicité externes. Ils ont été comparés par étude de simulation selon différents scénarios d'efficacité et de toxicité.

Résultats : Toutes les stratégies de schéma contrôlent le risque d'erreur de type I. Les schémas avec modélisation conjointe multinomiale de l'efficacité et de la toxicité ont une puissance supérieure. Les méthodes pondérées sur la fraction de suivi disponible lors des analyses intermédiaires permettent l'inclusion en continu des patients et ainsi une durée d'essai réduite en moyenne. La corrélation entre efficacité et toxicité influence les performances des schémas avec une augmentation du risque d'erreur de type I lorsque la corrélation est plus faible que planifiée.

Discussion : L'approche bayésienne avec modélisation multinomiale de l'efficacité et de la toxicité, avec pondération sur la fraction de suivi disponible lors des analyses intermédiaires semblent la plus avantageuse sur le plan des propriétés statistiques et de la durée de l'essai. De plus elle permet l'incorporation d'information extérieure via les distributions *a priori* utilisées pour l'inférence.

Mot-clefs : essais cliniques, phase II, schéma adaptatif, bayésien, cancer, immunothérapie.

TITLE : Comparison of phase II trial designs for the evaluation of new therapeutics in oncology

Introduction : Novel anti-cancer therapeutics like immunotherapies and target therapies have been challenging conventional clinical trial designs across all drugs development phases because of their different behaviour compared with cytotoxic chemotherapies. As a motivating example, we used a phase II trial evaluating a combination of new therapeutics in advanced oesogastric adenocarcinoma to exemplify the planning of an adaptive design in these settings : a prolonged observation window and sequential joint evaluation of efficacy and toxicity.

Methods : We defined 6 candidate adaptive designs combining the following approaches : Bayesian design allowing continuous accrual or not, separate monitoring of toxicity on top of efficacy or joint evaluation of both criteria with multinomial modelling, Simon's 2 stage design, use of external data. A simulation study under different scenarios of efficacy and toxicity was performed to compare the operating characteristics of the different designs.

Results : All designs controlled the type I error rate. Multinomial modelling approaches showed increased power. Methods that included weighting of observations according to the fraction of available follow-up at interim analyses resulted in a reduced trial duration by allowing continuous accrual. Correlation between efficacy and toxicity could influence designs performances by increasing the type I error rate when the correlation was weaker than expected.

Discussion : The Bayesian approach with multinomial modelling of efficacy and toxicity and weighting according to the fraction of available follow-up at interim analyses seems the most advantageous solution in terms of operating characteristics. Moreover, this approach allows adding external information via prior distributions.

Mot-clefs : clinical trials, phase II, adaptive design, bayesian, cancer, immunotherapy.