

ÉVALUATION DE DIFFÉRENTS DESIGNS D'ESSAIS CLINIQUES DANS LES CANCERS RARES: ÉTUDE DE SIMULATION

Mohamed Amine BAYAR ^{1 2 4} & Marie-Cécile LE DELEY ^{2 3 5}

¹ *École Supérieure de la Statistique et de l'Analyse de l'Information, 6, Rue des Métiers
- Charguia II - B.P 675 - 1080 Tunis*

² *Institut Gustave Roussy, 114, Rue Édouard Vaillant, 94805, Villejuif, FRANCE*

³ *Université Paris Sud, Faculté de Médecine Paris Sud, 63, rue Gabriel Péri, 94276, Le
Kremlin-Bicêtre Cedex*

⁴ *Mohamedamine.BAYAR@igr.fr*

⁵ *Marie-Cecile.LEDELEY@gustaveroussy.fr*

Résumé.

Problématique. L'essai clinique randomisé représente aujourd'hui le gold standard pour comparer un nouveau traitement à un traitement de référence ou à un Placebo. Le design classique de ce type d'essais suppose le recrutement d'un grand nombre de patients pour assurer un bon niveau de preuve communément défini par une erreur de première espèce d'au plus 5% pour un test bilatéral et par une puissance d'au moins 80% pour une taille d'effet souhaitée. Dans le cadre d'une maladie rare, les patients pouvant être inclus dans un essai clinique sont en nombre très limité. S'en tenir alors au design classique très exigeant en nombre de patients semble impossible et peut s'avérer contre-productif.

Objectif. Le but de ce travail de simulation est d'évaluer, sur une grande échelle de temps, la performance de différents designs d'essais cliniques afin d'identifier lesquels optimisent l'utilisation des ressources réduites dans le cadre des maladies à faible incidence, et en particulier, les cancers rares. Il ne s'agit donc pas d'évaluer un seul essai isolé à travers son niveau de preuve et sa puissance, mais d'estimer le bénéfice et le risque à plus long terme en considérant une série d'essais consécutifs.

Résultats. Réaliser des essais plus nombreux mais de plus petite taille, et utiliser des seuils d'erreur de première espèce relâchés allant de 0.1 à 0.2, peut améliorer le bénéfice attendu d'une série d'essais consécutifs sur les pathologies rares, tout en maintenant le risque de sélectionner à terme un traitement moins efficace que le premier traitement contrôlé à un niveau acceptable.

Mots-clés. essai clinique randomisé, maladies rares, design optimal.

Abstract.

The randomized controlled trial is admitted, today, to be the gold standard method for comparing a new treatment to a standard treatment or a placebo. But, in the setting of

low-incidence diseases and particularly rare cancers, using type I error typical level of 5 per cent for bilateral formulation and power level of 80 per cent requires a sample size so large that the trial becomes simply unfeasible, either because the number of patients that can be included in the trial is insufficient or because the accrual duration needed to recruit such a number is too long. Consequently, a compromise between type I error level, type II error for an expected treatment effect, sample size and trial duration should be found. Our simulation study concludes that using, as a measure of evidence, type I error levels larger than the typical 5 per cent (10 to 20 percent) for judging treatment differences and performing smaller but more numerous trials yields on average to larger survival gains over a long term research horizon than sticking to the traditional type I error of 5 per cent. Yet, relaxing the type I error level increases the risk of worsening the treatment efficacy at the end of the research, and so the trial design should be chosen so that it maintains this risk at a reasonable level.

Keywords. randomized controlled trail, low-incidence diseases, optimal design . . .

1 Problématique

Les essais cliniques de phase III constituent une étape expérimentale essentielle qui précède l'adoption ou le rejet d'un nouveau traitement. Il s'en suit que le niveau de preuve exigé par les autorités de santé pour autoriser un nouveau traitement est élevé, ce qui est très raisonnable dans le principe. Les essais cliniques de phase III doivent donc être de taille assez grande pour détecter une différence, si elle existe, entre les différents traitements testés en garantissant le niveau de preuve requis. Il existe cependant des circonstances où il est difficile voire impossible que l'on réussisse à recruter le nombre de sujets nécessaire vu la faible incidence de la maladie.

Un cancer est défini comme rare si sa prévalence est inférieure à 50/100 000 habitants et/ou son incidence est inférieure à 6/100 000/an. 20% des cancers tous types confondus sont des cancers rares, 100% des cancers chez l'enfant sont des cancers rares. De plus, avec le développement des connaissances quant à la biologie des cancers, l'essor de la médecine personnalisée et des thérapies ciblées, mêmes les cancers les plus fréquents sont aujourd'hui subdivisés en sous-groupes de cancers rares.

Dans le cadre des essais sur les cancers rares, les essais sont fréquemment réalisés à travers des collaborations nationales et internationales, souvent complexes. Mais l'élargissement géographique peut ne pas suffire à atteindre la taille d'échantillon souhaitée et la durée des essais est souvent tellement longue qu'au moment de la présentation des résultats, les traitements risquent de ne plus avoir aucun intérêt car ils seront dépassés. De plus, engager toutes les ressources humaines, financières et matérielles dans une seule direction de recherche et sur une grande période de temps limite l'opportunité de tester d'autres traitements éventuellement plus prometteurs. Ceci est particulièrement problématique dans le contexte actuel où le développement thérapeutique s'intensifie avec l'apparition

chaque année de très nombreux nouveaux traitements à évaluer.

Une des solutions souvent envisagée consiste à mener un essai clinique non randomisé, beaucoup plus économe en nombre de patients. Cependant, les conclusions de ce genre d'étude sans groupe contrôle randomisé risquent d'être fausses en raison des biais de comparaison à une valeur théorique ou à un groupe contrôle historique.

Quand le choix se porte sur un essai randomisé, malgré le nombre limité de patients éligibles, il est habituel de conserver le niveau de preuve de 5% bilatéral, quitte à accepter un risque de seconde espèce élevé et donc à mener un essai de moindre puissance qui risque de ne pas identifier un nouveau traitement efficace.

Vu qu'une solution optimale garantissant le contrôle des différents risques d'erreur n'existe pas dans un délai de temps raisonnable, il est légitime de chercher le meilleur compromis possible. Le but premier de notre étude de simulation est donc de réexaminer cette problématique de l'erreur de première espèce, de la puissance, de la taille de l'échantillon et de la durée de l'essai, en prenant un nouvel angle de vue qui considère l'essai clinique comme faisant partie d'une longue série d'essais consécutifs, et non comme une étape isolée du processus d'amélioration des traitements.

2 Description du cadre de simulation

L'étude est basée sur la simulation d'une série d'essais randomisés successifs indépendants, réalisés chacun sur deux groupes parallèles, de taille fixe, sans analyse intermédiaire. Chaque essai de la série évalue un nouveau traitement expérimental E contre un traitement contrôle C , ce dernier correspondant au traitement retenu au terme de l'essai précédent. Le critère de jugement étant à chaque fois la survie globale.

En considérant une période de recherche de 15 ans, nous avons fait varier un premier paramètre de design, qui est le nombre d'essais successifs réalisés au cours de la série, noté K . Étant donné que le taux de recrutement, noté η , et la période de suivi minimum, notée FU , sont fixés en amont de l'étude, le nombre de patients inclus dans chaque essai, noté N , découle du nombre d'essais successifs selon l'équation 1:

$$N = \left(\frac{15}{K} - FU\right).\eta \quad (1)$$

Le deuxième paramètre de design variable est le niveau de preuve caractérisé par l'erreur de première espèce, noté α , exigé pour adopter le traitement expérimental évalué comme nouveau traitement de référence. À ce stade de la simulation, le risque de deuxième espèce n'est pas contrôlé.

Les différents designs d'essais cliniques caractérisés par une allocation particulière des deux paramètres variables précédemment définis (nombre d'essais successifs K et risque de première espèce α) seront évalués pour différentes situations cliniques. Ainsi, nous avons considéré des pathologies plus ou moins fréquentes, la fréquence de la pathologie étant caractérisée par le taux de recrutement mais aussi des pathologies plus ou moins

graves. Étant donné que nous avons choisi un modèle exponentiel de survie, la gravité de la pathologie peut être caractérisée par la médiane de survie ou de manière équivalente par le taux de décès, noté λ .

L'effet des traitements futurs est caractérisé, sous l'hypothèse des risques proportionnels, par le taux de décès, λ^E , dans le bras expérimental E . L'effet des traitements futurs n'est jamais connu d'avance, nous avons donc considérés différentes évolutions du taux de décès à travers le temps qui nous semblent réalistes vu l'évolution des traitements au cours des dernières années et en se basant sur une étude de littérature. Nous avons aussi considéré des évolutions pessimistes de l'effet des traitements futurs afin de tester la robustesse de nos résultats et conclusions face à de telles hypothèses.

Afin d'assurer une convergence satisfaisante à nos résultats, nous avons effectué 5000 répétitions de chaque série d'essais conçue chacune selon un design particulier. L'ensemble de la simulation est effectué à l'aide du logiciel R.

3 Métriques

Afin d'évaluer et de comparer la performance des différents designs d'essais cliniques, nous avons proposé deux métriques. Elles ont pour but de calculer le bénéfice et le préjudice qui découlent de l'adoption d'un design particulier.

3.1 Hazard Ratio total au terme des 15 années de recherche

Le rapport de risque total noté \mathcal{HR}_{tot} est le rapport du taux de décès chez les patients qui seraient traités après les 15 ans de recherche par le traitement sélectionné suite au dernier essai, par le taux de décès chez les patients qui ont reçu le premier traitement de référence. Le rapport de risque total s'écrit:

$$\mathcal{HR}_{tot} = \frac{\lambda_{K+1}^C}{\lambda_1^C} \quad (2)$$

Le rapport de risque total peut de façon équivalente être traduit en termes de bénéfice relatif de survie. Le bénéfice relatif de survie noté BRS est défini par:

$$BRS = \left[\frac{1}{\mathcal{HR}_{tot}} - 1 \right] \quad (3)$$

Ainsi si le traitement sélectionné au terme des 15 ans permet d'obtenir une médiane de survie de 18 mois alors qu'elle était initialement de 12 mois, ce progrès peut se traduire par un rapport de risque total $\mathcal{HR}_{tot} = 0.67$ ou bien par un bénéfice relatif de survie $BRS = 50\%$.

3.2 Espérance du bénéfice total de survie au terme des 15 années de recherche

L'espérance du bénéfice relatif de survie au terme des 15 années de recherche est donnée par la formule suivante:

$$\mathbb{E}[BRS] = \mathbb{E} \left[\frac{1}{\mathcal{HR}_{tot}} - 1 \right] \quad (4)$$

3.3 Probabilité d'avoir choisi au terme des 15 années de recherche un traitement moins efficace que le premier traitement de référence

La probabilité d'avoir choisi au terme des 15 années de recherche un traitement moins efficace que le premier traitement de référence est donnée par la formule suivante:

$$Risque = \mathbb{P}[\mathcal{HR}_{tot} > 1] \quad (5)$$

Nous avons défini comme acceptable un design garantissant une probabilité inférieure à 0.02 pour ce risque.

4 Principaux résultats

Dans le cadre des études cliniques sur les maladies rares où le nombre de patients pouvant être inclus dans un essai est limité, le design classique qui consiste à effectuer des essais de grande taille pour atteindre le niveau d'erreur de première espèce de 5% ne constitue pas le design idéal lorsque le but est de maximiser le bénéfice de recherche à long terme. Notre étude montre clairement que faire le choix d'effectuer des essais de plus petite taille et supérieurs en nombre pourrait constituer un choix plus judicieux qui assure une espérance de bénéfice plus grande au terme des 15 ans de recherche.

Cependant, l'augmentation du nombre d'essais avec en corollaire la réduction de la taille de chaque essai et le choix d'un niveau d'erreur de première espèce supérieur à 5% sont conjointement limités par le risque de mal choisir les traitements qui ne doit pas dépasser une valeur seuil (c.f table 1). Les différentes évolutions à travers le temps du taux de décès, λ^E , dans le bras expérimental, sont notées par ordre de la plus optimiste à la plus pessimiste $E1$, $E2$ et $E3$.

Nous avons testé ces recommandations en supposant des hypothèses très pessimistes quant à l'effet des traitements futurs, elles restent acceptables. Nous avons aussi comparé nos résultats à ceux d'autres études qui adopte la même approche de maximisation de gain de survie sur une longue période. Nos résultats sont similaires aux résultats des autres études.

	Caractéristiques de la distribution		Design traditionnel $k = 2$ $n = 650$ $\alpha = 0.025$		Design optimal				
	$\mathbb{E}[\mathcal{HR}]$	$\mathbb{P}[\mathcal{HR} < 0.5]$	$\mathbb{E}[BRS]$	Risque	\hat{K}	\hat{n}	$\hat{\alpha}$	$\mathbb{E}[BRS]$	Risque
<i>E1</i>	0.950	0.02	26	0.0010	8	88	0.2	48	0.0072
<i>E2</i>	0.950	0.01	23	0.0018	8	88	0.2	42	0.0110
<i>E3</i>	0.975	0.02	22	0.0028	7	114	0.2	41	0.0198

Table 1: Définition du design optimal et comparaison à un design traditionnel pour trois hypothèses concernant l'effet du traitement

Bibliographie

- [1] Le Deley, M.C et Ballman, K.V et Marandet, J. et Sargent. D (2012), *Taking the long view: how to design a series of phase III trials to maximize cumulative therapeutic benefit*, Clinical Trials.
- [2] Sposto, R et Stram, D.O (1999), *A strategic view of randomized trial designs in low-incidence paediatric cancer*, Stat Med.
- [3] Djulbegovic, B et Kumar, A. et Glasziou, P.P. et Pereva, R. et Reljic, T. et Dent, L. et Raftery, J. et Johansen, M. et Di Tanna, G.L. et Miladinovic, B. et Soares, H.P et Vist, G.E. et Chalmers, I. (2012), *New treatments compared to established treatments in randomized trials*, Cochrane Database Syst Rev.