



Société Française de  
Pharmacologie et de Thérapeutique

Groupe de Travail Méthodologie

## Livre blanc SFPT

De la nécessité de la méthodologie  
dans l'évaluation des médicaments

*Document compagnon*

Dossier 12 – Etudes monobras  
et essais à contrôle externe

V2, 21 mars 2022



## Groupe de rédaction / relecture (Par ordre alphabétique)

- Theodora Angoulvant
- Laurent Bertolotti
- Jean-Luc Cracowski
- Michel Cucherat
- Dominique Deplanque
- Guillaume Grenet
- François Gueyffier
- Behrouz Kassai
- Charles Khouri
- Silvy Laporte
- Bruno Laviolle
- Jean-Christophe Lega
- Clara Locher
- Florian Naudet
- Edouard Ollier
- Antoine Pariente
- Matthieu Roustit
- Tabassome Simon



### [Licence Creative Commons](#)

Cette œuvre est mise à disposition selon les termes de la Licence Creative Commons Attribution 4.0 International

Vous êtes autorisé à :

- Partager — copier, distribuer et communiquer le matériel par tous moyens et sous tous formats
- Adapter — remixer, transformer et créer à partir du matériel pour toute utilisation, y compris commerciale.



# Table des matières

1	Introduction.....	7
2	Problématiques méthodologiques des études mono-bras.....	8
2.1	Des exemples récents de la « règle des 0%/100% » .....	9
3	Des études mono-bras aux essais à contrôle externe.....	13
4	Problématiques méthodologiques des comparaisons externes.....	14
5	Comparaison post hoc.....	14
6	Biais de confusion.....	14
7	Autres biais.....	15
8	Pertinence clinique.....	15
9	Étude de cas .....	16
10	Les conditions d'acceptabilité des comparaisons externes .....	17
11	La comparaison externe doit être formalisée .....	17
12	La comparaison externe doit être clairement explicitée .....	17
13	Il doit être possible d'écarter un choix arbitraire de la référence de comparaison, destiné à favoriser le traitement évalué.....	18
14	Les ajustements effectués doivent permettre d'écarter un biais de confusion. ....	19
15	La référence de comparaison doit être cliniquement pertinente et loyale.....	19
16	Les revues systématiques doivent être de bonne qualité.....	20
17	L'exposition potentielle aux biais de l'étude mono-bras et des études de références doit être acceptable .....	20
18	Le résultat suggéré par la comparaison externe doit être cliniquement pertinent.....	22
19	Synthèses des problématiques et de leurs solutions.....	23
20	Méta-recherche.....	25
20.1	Fiabilité des comparaisons externes .....	25
20.2	Etudes monobras et enregistrement FDA.....	26
21	Conclusion .....	27



# 1 Introduction

Les études mono-bras (single-arm study) consistent en une série prospective de patients recevant tous le traitement évalué. Elles donnent uniquement la valeur du critère de jugement sous traitement, « dans l'absolu » par exemple la valeur d'un taux de succès, de réponse, de survie, ou la valeur d'un paramètre biologique (« 42% de décès à 12 mois » par exemple). Ces études ne permettent donc pas d'apporter directement la démonstration du bénéfice d'un nouveau traitement. Pourtant, de plus en plus d'études mono-bras sont proposées comme seule étude « pivot » pour l'enregistrement, le remboursement ou la modification des pratiques. [1] [2].

## 2 Problématiques méthodologiques des études mono-bras

Les études mono-bras n'apportent, en fait, que la moitié de l'information nécessaire pour déterminer l'effet d'un traitement : une valeur dans « l'absolu » du critère de jugement chez les patients traités. Suivant les termes de l'épidémiologie moderne [3], il manque le contrefait (« counterfact ») qui permettrait de déterminer l'effet du traitement. La valeur sous traitement obtenue dans l'étude mono-bras est le « fait ». Pour vérifier s'il y a bien un effet du traitement, il faut le « contrefait ». Dans un essai randomisé, le « fait » est apporté par le groupe actif expérimental et le « contrefait » est apporté par le groupe contrôle expérimental (contrôle interne).

Pour déduire si cette valeur représente un bénéfice du traitement, il est nécessaire de procéder à une comparaison pour montrer que cette valeur est différente de celle qui aurait été obtenue chez ces patients, sans traitement (avec le traitement standard ou sous placebo). Dans le cas des essais mono-bras, cette comparaison ne peut s'effectuer qu'avec des données externes à l'étude (contrôle historique par exemple). Cette comparaison externe est l'élément essentiel, à tel point que la littérature dans le domaine, comme ICH<sup>1</sup>, n'aborde pas cette problématique sous l'angle des études mono-bras, mais bien sous celui de la comparaison externe (*externally controlled trial*) [4].

Problématique méthodologique spécifique (Exposant à un risque de production de résultat favorable à tort au traitement étudié)	Démonstration que doivent apporter les solutions à ces problématiques (pour garantir la disparition du risque de conclure à tort)
Impossibilité de raisonnement contrefactuel, impossibilité de déterminer l'effet du traitement	Utilisation d'un groupe contrôle externe avec une comparaison indirecte non ancrée (pour prendre en compte les biais de confusion et autres). Réalisation d'une étude comparative contrôle externe (cf. section suivante)

Les études mono-bras ne permettent de conclure à l'intérêt du traitement évalué à elles seules que dans des situations exceptionnelles dites 0%/100% correspondant à l'observation confirmée d'évolutions favorables avec le nouveau traitement dans une pathologie où l'évolution serait péjorative dans 100% des cas (comme le traitement insulino-dépendant). Ce type de situation est exceptionnel<sup>2</sup> [5] (cf. section Des exemples récents de la « règle des 0%/100%»). Même dans le cas de l'infection par le virus Ebola par exemple, le pronostic, bien que très péjoratif, n'est pas suffisamment sombre pour rentrer dans la règle des 0%/100% avec l'existence de patients qui survivent spontanément. Un essai randomisé a donc été réalisé pour évaluer l'efficacité clinique du ZMapp [5].

<sup>1</sup> ICH E10, section 2.5 page 24.

<sup>2</sup> Par exemple même dans le cas de l'infection par le virus Ebola le pronostic, bien que très péjoratif, n'est pas suffisamment sombre avec l'existence de patient qui survive spontanément pour rentrer dans la règle des 100%/0%. Un essai randomisé a donc été réalisé pour évaluer l'efficacité clinique du ZMapp (7. Group PIW, Multi-National PIIST, Davey RT, Jr., Dodd L, Proschan MA, Neaton J, et al. A Randomized, Controlled Trial of ZMapp for Ebola Virus Infection. N Engl J Med. 2016;375(15):1448-56.).



De nombreux retours d'expérience donnent des exemples de traitements considérés comme apportant un bénéfice avec une comparaison non formalisée des résultats d'une étude mono-bras par rapport à une référence subjective ou implicite, mais pour lesquels ce bénéfice n'a pas été mis en évidence ultérieurement lorsque d'un ECR a été réalisé [6, 7] :

Traitement	Étude mono-bras	Phase 3 non concluante
Epacadostat par-dessus pembrolizumab dans le mélanome avancé	ECHO-202/KEYNOTE-037 (Taux de réponse 55%)	PFS et OS négatives ECHO-301/KEYNOTE-252 (taux de réponse groupe traité 34%)
Pembrolizumab monothérapie dans le cancer de la vessie	KEYNOTE-0523	KEYNOTE-361 arrêté prématurément pour surmortalité <sup>4</sup>
Pembrolizumab dans le carcinome hépatocellulaire	KEYNOTE-224	KEYNOTE-240 non conclusive <sup>5</sup>
L'atezolizumab en première ligne du cancer de la vessie métastatique	IMvigor 210	IMvigor 211 non concluante
Remdesivir dans la COVID-19	SIMPLE [8]	RECOVERY, DISCOVERY

La réalisation d'étude mono-bras est souvent justifiée par l'impossibilité de réaliser un essai comparatif randomisé en raison d'un nombre trop faible de patients. En oncologie, Rittberg et al. a montré que pour 31 enregistrements récents basés sur une étude mono-bras, un essai randomisé aurait pu réalisable dans plus de 80% des cas [9].

## 2.1 Des exemples récents de la « règle des 0%/100% »

Jusqu'à présent il existait peu d'exemples de la règle des 0/100% hormis des situations historiques comme les antibiotiques dans la méningite tuberculeuse par exemple. Les nouveaux traitements ne s'attaquaient pas à des situations pathologiques très avancées (la partie 100% de la règle) car il est difficile d'espérer en général un quelconque bénéfice dans ces situations. De même, il était aussi rare qu'un traitement apporte un bénéfice complet ou presque complet aux patients (la partie 0% de la règle).

Cependant, de telles situations commencent à apparaître, en particulier dans le champ de la thérapie génique, dont voici deux exemples.

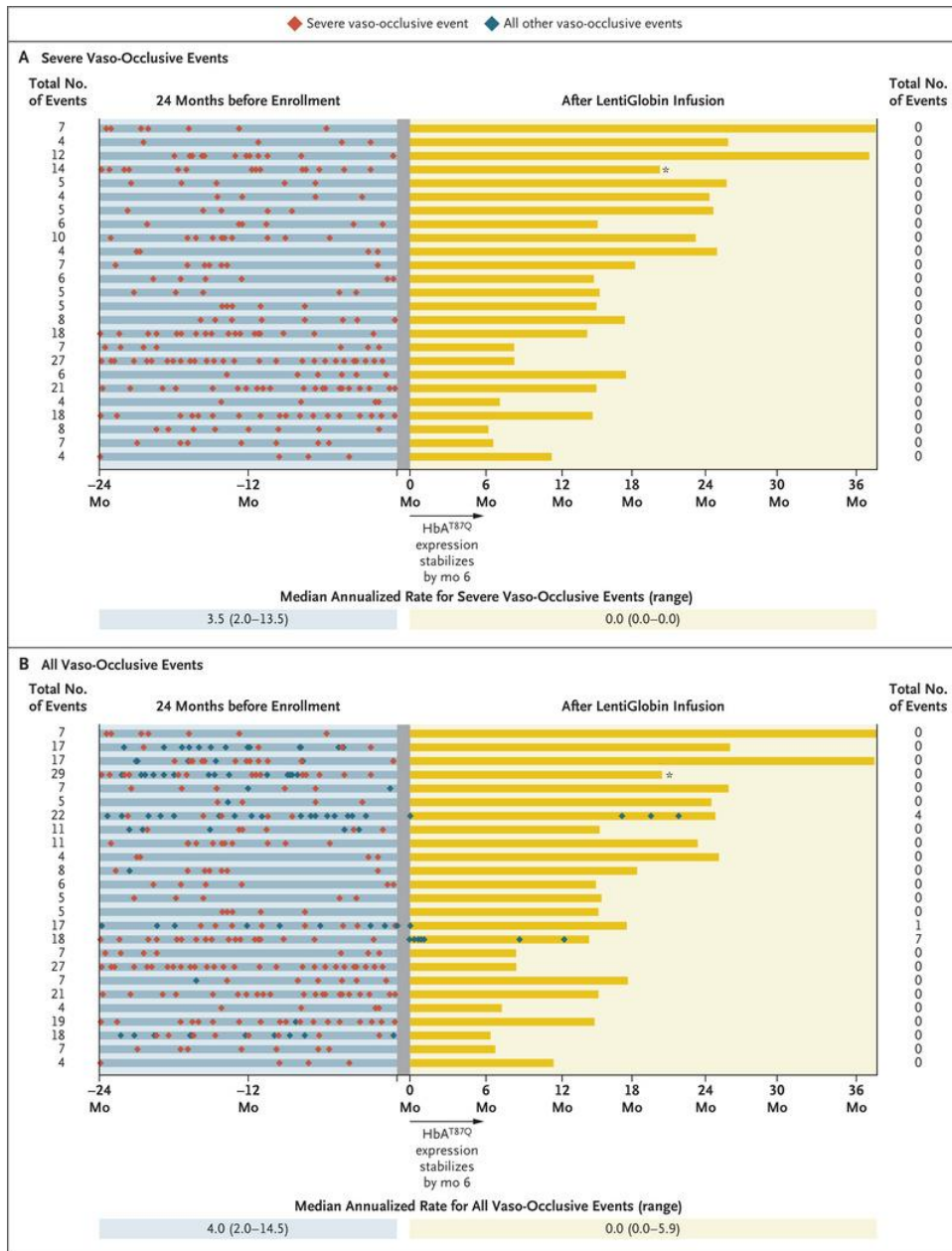
Le premier exemple est apporté par la thérapie génique lovetibeglogene autotemcel dans le traitement de la drépanocytose [10]. Cet essai monobras de 35 patients montre un changement avant-après dans la fréquence des crises vaso-occlusives d'une très grande ampleur, évident sur une simple représentation graphique (Figure 1). Le contrefait aux résultats obtenus chez les patients traités est fourni par la comparaison à la fréquence des crises chez les mêmes patients avant l'administration du traitement (comparaison avant– après<sup>6</sup> proche d'une analyse de séries temporelles interrompues).

3 10.1016/S1470-2045(17)30616-2

4 <https://www.esmo.org/Oncology-News/Patients-with-mUC-and-Low-PD-L1-Expression-May-Have-Decreased-Survival-When-Treated-with-Pembrolizumab-or-Atezolizumab>

5 <https://www.mrknewsroom.com/news-release/oncology/merck-provides-update-keynote-240-phase-3-study-keytruda-pembrolizumab-previous>

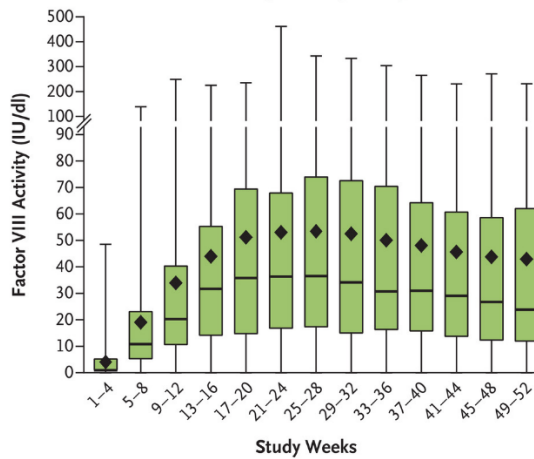
<sup>6</sup> Même s'il existe des réserves en inférence causale sur l'aptitude des comparaison avant-après en situation observationnelle à établir une vraie relation de causalité [11].



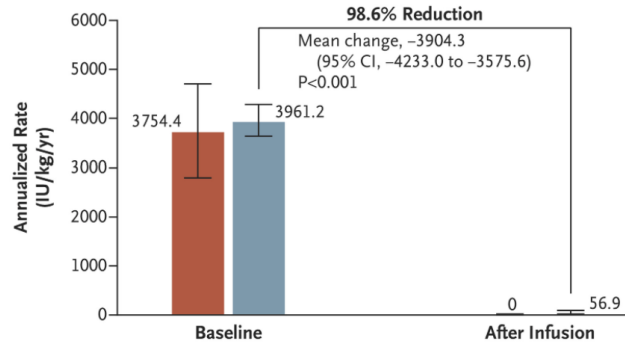
**Figure 1 – Représentation graphique du changement de fréquence des crises vaso-occlusives avant et après l’introduction du lovetibeglogène autotemcel dans la drépanocytose [10]**

Un autre exemple est fourni par le valoctogène roxaparavec dans l’hémophilie A [12]. Sur le critère intermédiaire du niveau d’activité du facteur VIII, un effet important de restitution de la fonction est observé, en rupture avec le niveau prétraitement. Un effet encore plus spectaculaire est observé sur les besoins en facteur VIII (Figure 2).

A Modified Intention-to-Treat Population (N=132)



A Factor VIII Use



**Figure 2 – Effet du valoctocogène roxaparvovec dans l’hémophilie A sur le niveau d’activité du facteur VIII (figure de gauche) et en besoin de facteur VIII (figure de droite) [12].**

Cependant il convient de noter la variabilité inter sujets, qui apparaît encore plus importante sur le critère clinique de la fréquence des saignements (Figure 3). À son niveau, la règle des 0%/100% n’est plus aussi évidente, avec l’existence de patients sans saignement avant le traitement (remise en cause du 100%) et d’autres dont la fréquence augmente après traitement (remise en cause du 0%) y compris chez des sujets sans saignement avant traitement. Ces motifs suggèrent une variabilité des évolutions temporelles du taux de saignement. Cette variabilité des évolutions pourrait conduire aux mêmes observations en l’absence de réel bénéfice du produit sur le critère clinique, sans qu’il puisse être possible de l’objectiver en raison de l’absence de groupe contrôlé.

Il apparaît ainsi que, même s’il semble possible, dans certains cas, de conclure à un effet biologique (pharmacologique) d’un produit en l’absence du contrefait, les résultats obtenus sur les critères cliniques peuvent être moins clairs et nécessiter un groupe contrôlé pour s’assurer que l’effet biologique se traduit bien en bénéfice clinique pertinent. Il s’agit en fait de la classique problématique de la traduction des effets biologiques en bénéfice clinique (cf. livre blanc, section 2.1). En cas d’une variabilité importante inter patients du critère clinique, le recours à une comparaison devient alors indispensable pour démontrer le bénéfice clinique. Ce point est particulièrement prégnant avec les maladies rares où il est classique de noter un spectre important dans les formes cliniques et les évolutions. Du fait de la rareté des patients, de fortes variabilités inter patients du critère clinique seront alors présentes dans les études non comparatives, d’où l’importance de la documentation de l’histoire naturelle de la maladie [13]. Enfin, cette variabilité souligne aussi la pertinence de rapporter les résultats individuels dans ce type d’étude, comme cela a été fait dans les deux exemples précédemment cités (Figure 1 et Figure 3).

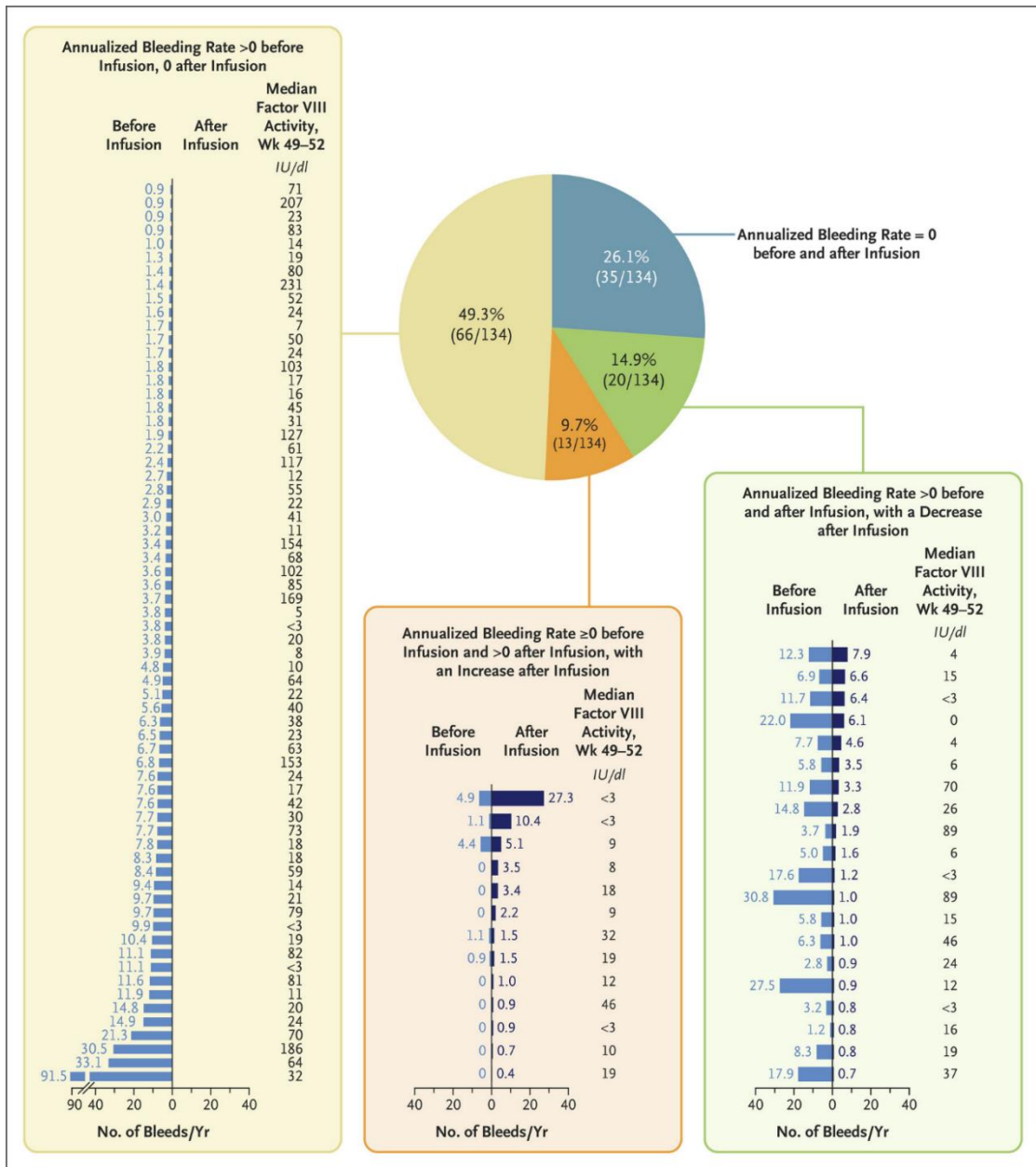


Figure 3 – Effet avant après sur la fréquence des saignements du valoctocogene roxaparvovec dans l'hémophilie A [10].

### 3 Des études mono-bras aux essais à contrôle externe

Les études mono-bras ne permettent pas d'évaluer le bénéfice clinique en raison de l'absence de raisonnement contrefactuel.

Dans les rares situations où ces études seraient proposées, elles doivent obligatoirement prendre la forme d'études comparatives à contrôles externes

Les études à contrôle externe sont des études qui formalisent la comparaison des résultats d'une étude mono-bras avec celle d'un contrôle externe (contrôle historique par exemple). Ces études peuvent être prévues *a priori*, au moment de la planification de l'étude mono-bras, mais le plus souvent, elles sont réalisées *a posteriori* après la disponibilité des résultats de l'étude mono-bras.

*Dans les études mono-bras (cf. section 2), en l'absence de point de comparaison apporté par un groupe contrôle, il est impossible de savoir si la valeur du critère de jugement observée sous traitement est meilleure que celle qui aurait été obtenue sans ce traitement dans la même étude (mêmes patients, même mesure du critère de jugement, même suivi, même contexte de soins). L'exploitation des études mono-bras pour démontrer le bénéfice d'un nouveau traitement nécessite donc une comparaison implicite qui fera obligatoirement appel à une référence de comparaison externe (comme une comparaison historique par exemple).*

Plusieurs types de comparateurs externes sont envisageables (cf. Tableau 1).

**Tableau 1 – Types de comparateurs externes possibles**

- Norme, valeur de référence qui peut être soit fixée par une « exigence réglementaire », soit issue de la littérature (étude de cohorte par exemple) ou d'une conférence de consensus
- Cohorte disponible sous la forme d'une publication (données individuelles inaccessibles) ou publication d'un essai randomisé sont un des groupes peut servir de comparaison externe qui pourra servir de base à une MAIC ou d'une autre méthode de ce type
- Cohorte avec accès aux données individuelles qui pourra servir à une approche type étude observationnelle ou à la constitution d'un groupe contrôle synthétique
- Données individuelles d'un essai randomisé qui pourra servir à une approche type étude observationnelle
- Évolution des mêmes patients avant le traitement (design autocontrôlé de type avant/ après)
- Emprunt de données (voire document de synthèse sur les nouvelles méthodologies, section 19)

Les comparaisons externes sont souvent appelées historiques car les groupes contrôles servant à la comparaison sont non contemporains de l'étude monobras.

## 4 Problématiques méthodologiques des comparaisons externes

Les comparaisons externes présentent plusieurs limites méthodologiques fortes.

### 5 Comparaison post hoc

La première est que, le plus souvent, le choix de la référence de comparaison s'effectue de manière post hoc (non prévue au protocole, souvent après que les résultats de l'étude monobras soient connues) et que rien ne permet de garantir que ce choix n'ait pas été fait pour favoriser le nouveau traitement. La solution à cette limite est simple. Ces études ne doivent plus être conçues comme des essais mono-bras descriptif, mais comme de véritables essais comparatifs dont le groupe contrôle n'est pas interne et contemporain, issu d'une randomisation, mais externe (Tableau 2). De ce fait l'objectif de l'étude est bien celui de montrer l'intérêt clinique du nouveau traitement (et non pas de décrire l'évolution des patients avec le traitement). Le choix de la référence de comparaison est fixé au protocole ainsi que la méthode d'analyse, empêchant tous choix post hoc. Cette approche est avalisée dans la nouvelle version de ICH E10 qui propose le terme d'essai à contrôle externe (*externally controlled trial*) [14].

La détermination *a priori* de la référence de comparaison au moment de la construction du protocole de l'étude permet aussi une meilleure prévention des biais en utilisant, par exemple, le même critère de jugement et la méthode de sa mesure (cf. infra).

**Tableau 2 – Apport d'une véritable étude comparative à contrôle externe par rapport aux études mono-bras telles que réalisées actuellement.**

	<b>Étude mono-bras</b> (telle que réalisée actuellement)	<b>Étude comparative à contrôle externe</b>
Objectif	Décrire le devenir des patients avec le nouveau traitement	Montrer la supériorité du nouveau traitement
Détermination du bénéfice du traitement	En post hoc, par une analyse complémentaire souvent mise en place après l'obtention des résultats	Objectif de l'essai défini a priori dans le protocole
Choix de la référence de comparaison	Post hoc, potentiellement influencée par les résultats observés de l'étude	Fixée <i>a priori</i> , prévue par le protocole

### 6 Biais de confusion

L'autre limite majeure de la comparaison à un contrôle externe est le biais de confusion : rien ne garantit que les groupes ou les résultats comparés sont réellement comparable.

Plusieurs méthodes de formalisation des comparaisons externes (aussi appelée parfois comparaison indirecte, *unanchored indirect comparison*) ont été proposées afin de tenter de fiabiliser les conclusions tirées des études à comparateur externe. Ces approches ont pour objectif de suppléer

l'absence de contrôle par design (plan expérimental) des facteurs de confusion par leur prise en compte dans l'analyse statistique (ajustement). Les méthodes permettent en théorie de supprimer le biais de confusion à condition que tous les facteurs de confusion puissent être pris en compte (facteurs conditionnant le critère de jugement et ayant une distribution différente entre la groupe traitée et le comparateur externe, y compris les modificateurs de l'effet du traitement). Pour cela les facteurs de confusion potentiels de chaque critère de jugement doivent être préalablement identifiés par une revue systématique des études pronostiques. L'ajustement statistique réalisé (quelle que soit la méthode) devra prendre en compte tous ces facteurs. Pour confirmer que ces ajustements ont permis de supprimer le biais de confusion, l'importance du biais de confusion résiduel doit être appréciée par l'utilisation de contrôles négatifs comme variables de falsification et par une analyse quantitative du biais (« *bias analysis* ») [8]. Compte tenu des difficultés rencontrées pour apporter ces garanties, les résultats ne seront convaincants qu'en cas de résultats montrant un large effet, non explicables par les biais et la confusion résiduelle.

Le choix *a priori* de la référence de comparaison au moment de la construction du protocole contribuera aussi à la minimisation du biais de confusion en permettant d'aligner les critères d'éligibilité avec ceux utilisés pour la construction du contrôle externe (approche par restriction).

Cette limite du biais de confusion fait qu'au niveau réglementaire, ICH E10 n'envisage le recours aux essais à contrôles externes que dans des situations très particulières où l'effet du traitement est extrêmement important et le cours de la maladie hautement prévisible (« *Use of the external control design is restricted to situations in which the effect of treatment is dramatic and the usual course of the disease highly predictable* ») [14].

## 7 Autres biais

D'autres limites existent en termes de biais de mesure, de sélection et d'attrition. En particulier en termes de biais temps-dépendants et de synchronisation des débuts de suivi [15].

## 8 Pertinence clinique

L'étude à comparateur externe doit aussi assurer la pertinence clinique de la comparaison effectuée en termes de : critères de jugement principaux pertinents, comparateur loyal représentant le traitement recommandé « *standard of care* » en vigueur au moment de la prise de décision, durée de suivi, évaluation de la sécurité. Sur ces points, l'étude doit être comparable à ce qu'aurait été un essai randomisé pivot.

## 9 Étude de cas

L'atezolizumab a bénéficié d'un enregistrement accéléré dans le traitement de première ligne du cancer de la vessie métastatique de patients non éligibles à un traitement par platine sur la base de l'étude mono-bras IMvigor 210 [16] qui montrait un taux de réponse de 23%, statistiquement significatif par rapport à une valeur de référence de 10% fixée au protocole. La publication ne donne pas de justification factuelle de la norme, qui est présentée comme étant le taux de réponse attendue chez ces patients compte tenu des traitements utilisés.

	Patients	Complete response	Partial response	Objective response, n (% [95% CI])*	Median duration of response (95% CI)
	119	11	16	27 (23% [16-31])	NE (14-1-NE)
IC2/3	32	4	5	9 (28% [14-47])	NE (11-1-NE)
IC1/2/3	80	8	11	19 (24% [15-35])	NE (NE-NE)
IC1	48	4	6	10 (21% [10-35])	NE (NE-NE)
IC0	39	3	5	8 (21% [9-36])	NE (12-8-NE)

Data cutoff was July 4, 2016. PD-L1=programmed death-ligand 1. IC=tumour-infiltrating immune cell. NE=not estimable. \*Includes objective response rate per Response Evaluation Criteria in Solid Tumors version 1.1 (independent review facility).

Table 2: Objective response by PD-L1 status on tumour-infiltrating immune cells

Cet enregistrement accéléré était conditionné à la réalisation de la phase 3 comparative versus chimiothérapie (IMvigor 211) [17]. Cet essai a été négatif sur son critère de jugement principal, ne montrant pas de bénéfice sur la survie globale (HR 0.87 [0.63 ; 1.21]).

De plus, le taux de réponse dans le groupe chimiothérapie a été de 21.6%, très proche des 23% du critère de jugement principal de l'étude mono-bras IMvigor210, montrant bien, a posteriori, le caractère arbitraire et subjectif de fixer le taux de réponse attendue sous chimiothérapie à 10%.

Cet exemple illustre la difficulté de déterminer la valeur contrefactuelle (« *counterfactual value* ») du critère de jugement, c'est-à-dire ce qui aurait été observé, chez ces patients, sans le nouveau traitement (et qui aurait donné par le groupe contrôle d'un essai comparatif). Lorsqu'une valeur est précisée (comme ici 10%), cette valeur peut être arbitraire, basée sur un avis d'expert qui est une évaluation subjective de cette valeur contrefactuelle. Compte des enjeux de ces études la crainte est qu'une valeur particulière favorable au traitement étudié soit choisie (comme ce fût le cas dans cette étude). Le but des essais à contrôle externe de bonne méthodologie est d'éviter de tels choix arbitraires.

On peut aussi noter que dans cette étude le critère de jugement était de faible pertinence clinique et ne correspondait pas au critère attendu, l'OS, (qui a bien été pris en compte comme critère de jugement principal dans l'essai subséquent de phase 3). Ces études conduisent donc à ce que la prise de décision d'utiliser le nouveau traitement s'effectue non seulement en se basant sur une méthodologie très dégradée, mais aussi sur des données sans réelle pertinence clinique. Sans formalisation par une vraie comparaison externe l'utilisation isolée de ces études mono-bras s'apparente à un déni complet d'évaluation.



## 10 Les conditions d'acceptabilité des comparaisons externes

Les conditions d'acceptabilité des comparaisons externes pour positionner un nouveau traitement dans la stratégie thérapeutique dépendent des solutions apportées aux diverses problématiques méthodologiques que posent les études à contrôles externes (cf. section 4).

### 11 La comparaison externe doit être formalisée

La comparaison externe doit avoir été planifiée *a priori*, en même temps que l'étude mono-bras. L'étude devait avoir pour objectif de déterminer le bénéfice du traitement par comparaison externe et non pas simplement de décrire le devenir de quelques patients tous traités.

La méthode de comparaison externe doit être clairement définie dans les méthodes : comparaison à une norme, comparaison « ajustée » par rapport à une cohorte de référence (*Matching-adjusted indirect comparisons* (MAIC) [18, 19], « ajustement » traditionnel), simulation du taux de référence (*Simulated Treatment Comparison*), etc.). En d'autres termes, une étude non comparative purement descriptive<sup>7</sup> ne sert à rien pour l'évaluation du bénéfice d'un traitement. Pour être exploitable dans ce but, l'étude doit être une étude de comparaison externe et doit donc le prévoir explicitement dans son objectif et sa méthode. Ainsi ICH E10 (version de juillet 2000<sup>8</sup>) ne mentionne que les études à comparateur externe (section 2.5) et ne parle pas d'études mono-bras, non comparatives.

Une telle étude doit donc être mise en place prospectivement et prévue au plan de développement comme étant l'étude « finale » devant apporter les preuves cliniques de l'intérêt du traitement. Il s'agit donc d'une phase 3 stricto sensu. En effet rien n'empêche de prévoir dans un plan de développement que la démonstration ultime du bénéfice du traitement sera apportée par un essai non comparatif avec un comparateur externe (cf. ICH E10).

### 12 La comparaison externe doit être clairement explicitée

La référence de comparaison doit être une étude ou une valeur déduite d'une étude et non pas une valeur arbitraire issue d'un avis d'expert sans justification factuelle.

---

<sup>7</sup> Souvent le sponsor industriel de l'étude non comparative insiste sur cet objectif purement descriptif afin de prévenir toute utilisation de cette étude en défaveur du traitement si, par exemple, les résultats pouvaient être utilisés pour conclure à une faible ou à une non-efficacité ou à un surcroît d'effet indésirable. Il y a donc une reconnaissance de facto par le monde industriel de l'insuffisance de ces études avec cet objectif de conduire à une évaluation des effets des traitements. En fait ces études contribuent à un jeu de dupe : si elles donnent des valeurs peu flatteuses pour le traitement évalué, l'objectif purement descriptif est mis en avant pour prévenir toute exploitation des résultats en défaveur du traitement. En revanche, si les résultats peuvent être exploités en faveur du traitement, elles sont alors proposées comme étude d'évaluation du bénéfice, donc comme base à une comparaison externe. Tout cela de façon purement post hoc, uniquement en fonction des résultats obtenus. D'où l'importance du respect des objectifs initiaux des études

<sup>8</sup> <http://www.ich.org/products/guidelines/efficacy/efficacy-single/article/choice-of-control-group-and-related-issues-in-clinical-trials.html>

La base de la comparaison doit être clairement explicitée :

- Pour la comparaison à une norme, la valeur de norme retenue doit être justifiée à partir de données factuelles (études des traitements précédents, études d'histoire naturelle, etc.), nonobstant les difficultés à fixer une norme représentative des patients recrutés dans l'étude
- Pour les autres types de comparaisons externes (MAIC, etc.) : la ou les études de référence qui serviront de base à la comparaison

### 13 Il doit être possible d'écarter un choix arbitraire de la référence de comparaison, destiné à favoriser le traitement évalué

Il doit être possible d'écarter un choix post hoc de la base de la comparaison externe, drivé par la valeur obtenue par l'étude mono-bras, par un choix clairement effectué avant la date de prise de connaissance des résultats (analyse intermédiaires ou analyse finale).

Il doit être démontré que la référence finalement retenue n'est pas celle la plus favorable à la mise en évidence du bénéfice du traitement évalué. Fréquemment il existe plusieurs études pouvant servir de base à la comparaison externe. Une forte variabilité des résultats entre études est aussi fréquemment rencontrée. Il est donc possible de choisir, arbitrairement, parmi toutes les études disponibles, celle ayant donné les pires valeurs, ce qui in fine favorisera le plus le nouveau traitement. Ce problème existe aussi si ce choix est fait *a priori*, avant d'avoir les résultats de la mono-bras, en prenant le groupe contrôle avec les plus mauvais résultats.

La seule possibilité permettant d'exclure formellement une telle démarche est de disposer de la revue systématique de toutes les études pouvant potentiellement servir de base de comparaison. Ensuite, l'étude retenue doit être celle la plus défavorable au nouveau traitement, dans le but d'être conservateur. Une attitude conservatrice est indispensable afin d'apporter des garanties solides avec un design fragile : il doit donc être possible d'exclure tous les points faisant la fragilité de ces comparaisons externes, par exemple le choix d'un comparateur favorable.

Il est aussi possible que des analyses de sensibilité soient réalisées avec toutes les autres études pouvant servir potentiellement de référence et que toutes ces analyses de sensibilité montrent la supériorité du nouveau traitement.

La revue systématique des études de référence doit être loyale (cf. ci-dessous).

L'approche de constitution de **groupe contrôle synthétique** par sélection dans les bases historiques de patients similaires à ceux traités avec le nouveau traitement dans l'étude mono-bras expose à un risque important et non contrôlable de « data dredging ». En effet, cette approche consiste à extraire, dans une base contenant de nombreux cas, quelques patients bien sélectionnés similaires aux patients traités pour constituer un pseudo groupe contrôle. Il est facile de construire un processus de sélection testant toutes les possibilités de constitution de groupe contrôle de taille n qu'offre la base de données. Plusieurs bases peuvent être testées à la recherche d'un groupe contrôle optimal. L'acceptabilité de ce type d'approche n'est possible que si des garanties formelles d'absence de processus de data dredging de ce type sont apportées par un protocole strict, ne laissant aucune marge de manœuvre pour faire une sélection des contrôles drivée par les résultats.

## 14 Les ajustements effectués doivent permettre d'écarter un biais de confusion.

Pour prétendre avoir un niveau de preuve similaire à celui des essais randomisés, l'étude doit pouvoir apporter la preuve de l'absence d'un biais de confusion résiduel.

Pour cela, une revue systématique de bonne qualité doit avoir été entreprise pour déterminer cette liste des facteurs pronostiques du/des critères de jugements considérés (les facteurs pronostiques peuvent être différents en fonction du critère de jugement) des facteurs modifiant l'effet du traitement.

Ensuite, les ajustements statistiques réalisés doivent prendre en compte non seulement l'ensemble des facteurs pronostiques mais également les modificateurs d'effet.

Les modificateurs de l'effet doivent avoir été aussi inclus.

Pour prétendre avoir un niveau de preuve similaire à celui des essais randomisés, ces approches doivent apporter la preuve de l'absence de biais de confusion résiduelle.

Pour juger si cette condition de validée est vérifiée, il convient qu'une recherche préalable de tous les facteurs pronostiques et modifiants les effets a été conduite par des revues systématiques. Ensuite, il convient que les ajustements aient pris en compte tous ces facteurs. Le plus souvent cela est impossible en raison de l'indisponibilité de certaines de ces covariables (absence dans la publication de l'étude de référence pour une MAIC par exemple ou par absence de recueil dans l'étude).

Si un ou plusieurs de ces facteurs n'a pas pu être pris en compte, il convient que des analyses soient réalisées afin de quantifier par simulation le biais de confusion résiduelle (« *quantitative bias analysis* ») et de montrer que la différence obtenue après ajustement ne peut pas provenir d'un tel biais [20, 21, 22]. L'utilisation de contrôles négatifs est aussi possible. Dans tous les cas, une discussion approfondie du biais de confusion résiduelle soit être avoir été effectuée dans le rapport de l'étude en se basant sur les approches actuelles de quantification et de simulation.

Une partie du biais de confusion peut provenir de variables non liées aux patients mais liées au contexte des études : évolution séculaire et différence de risque intrinsèque des populations, des déterminants génétiques, des contextes de soins, des traitements concomitants, des stratégies de recours aux soins palliatifs, etc. Il s'agit de covariables pour lesquelles les ajustements ne peuvent pas être effectués. La référence de comparaison doit donc être le plus proche possible du contexte de soins de l'étude mono-bras.

La disponibilité d'une revue systématique est indispensable pour permettre de vérifier que tous les facteurs pronostiques ont bien été pris en compte. Une liste déclarative obtenue par avis d'experts est très insuffisante, car rien ne permet d'exclure que celle-ci n'ait pas été conditionnée, entre autres, par la disponibilité des variables dans les différentes études.

## 15 La référence de comparaison doit être cliniquement pertinente et loyale

Les patients utilisés comme référence de comparaison doivent avoir reçu les meilleurs traitements actuellement disponibles (« *standard of care* ») afin de représenter une base de comparaison loyale. La problématique est la même que celle concernant le choix du comparateur pour toutes études versus traitement actif.

Les patients doivent avoir reçu les traitements contemporains et ne peuvent pas être considérés comme sous-traités au regard des recommandations, pratiques et données actuelles.

Les groupes traités des essais randomisés de validation des traitements actuels apportent cette garantie. Ainsi ils doivent avoir été considérés de manière systématique.

Il est à noter que toutes les approches de formalisation des comparaisons externes peuvent aussi être utilisées pour comparer le nouveau traitement à plusieurs traitements de références (efficacité et sécurité relative), à la manière des méta-analyse en réseau [23].

## 16 Les revues systématiques doivent être de bonne qualité

Les revues systématiques entreprises pour identifier toutes les études de référence potentielles et lister les facteurs pronostiques du/des critères de jugement doivent être aux standards actuels (par exemple *Cochrane handbook*).

Une attention toute particulière sera apportée aux critères d'exclusion et à la liste des études exclues et des raisons de ces exclusions qui représentent le principal moyen de sélectionner les études favorisant le traitement étudié.

Pour prétendre à l'exhaustivité, la revue systématique doit être basée sur plusieurs bases bibliographiques (Pubmed plus au moins une autre). Il est à noter que le registre CENTRAL de la collaboration Cochrane, étant focalisé sur les essais randomisés, n'a qu'une très faible pertinence pour ce type de revue systématique.

La qualité méthodologique des études doit être prise en compte avec des outils d'évaluation du risque de biais approprié (comme PROBAST pour les études pronostiques(13))

## 17 L'exposition potentielle aux biais de l'étude mono-bras et des études de références doit être acceptable

Les méthodologies des 2 études doivent être comparables en termes de sélection, suivi des patients et mesure du critère de jugement pour exclure la possibilité d'une comparaison biaisée par des différences de méthode entre les études.

**La taille d'effet doit être très importante** pour qu'elle ne puisse pas provenir des biais (liés aux différences des méthodes entre les 2 études)

L'étude de comparaison externe doit discuter soigneusement tous les biais potentiels et apporter une argumentation convaincante comme quoi, quantitativement, les biais ne peuvent pas expliquer la taille de la différence.

La question des biais dans ce contexte est particulière, car il ne s'agit pas des biais traditionnels des études descriptives (aptitudes à estimer le paramètre de la population), mais bien des biais pouvant affecter la comparaison externe. Il convient donc de juger de la méthodologie des 2 études, l'étude

mono-bras et l'étude de référence, dans la perspective d'un biais dans la comparaison externe qui sera effectuée à partir de ces 2 études<sup>9</sup>.

Un **biais de mesure** surviendra quand la façon de mesurer le critère de jugement sera différente entre les 2 études, conduisant à une sous-estimation relative de la fréquence du critère de jugement dans l'étude mono-bras par rapport à l'étude de référence. Par exemple avec une définition des événements plus restrictive dans l'étude mono-bras que dans l'étude de référence ou une méthode de recherche des événements relativement plus sensible dans l'étude de référence ou relativement plus spécifique dans l'étude mono-bras. Une analyse comparative soignée des définitions et des méthodes est donc nécessaire. Cette description doit être discutée dans l'étude et toutes les analyses de sensibilité nécessaires réalisées. Une attention toute particulière doit être portée sur la définition des censures qui peuvent impacter les estimations de PFS par exemple [24, 25]. Finalement il n'y a guère que les critères purement objectifs (donc la mortalité totale) qui sont d'emblée à l'abri de ce type de biais.

**L'horizon d'analyse** (événements enregistrés uniquement durant la période de traitement ou durant tout le suivi) ainsi que la durée moyenne de suivi doivent aussi être comparables entre les études, car ces 2 paramètres influencent les fréquences absolues d'événements. L'utilisation d'un taux (densité d'incidence, en patients\*mois par exemple) solutionne les différences de durées de suivi moyennes, mais pas celui de l'horizon d'analyse.

Le **biais d'information** est en général facilement excluable, mais peut survenir au cours du temps si la dynamique d'arrêt du traitement est différente entre les 2 études, et ce indépendamment des conséquences de l'efficacité ou des effets indésirables sur la maintenance des traitements).

Un **biais d'attrition** est fréquemment possible comme avec une étude mono-bras adoptant une logique d'analyse per protocole comparée à une étude de référence dont l'objectif était purement descriptif avec une logique d'analyse en ITT. Par exemple, avec un dispositif ou un produit de thérapie cellulaire, l'analyse de l'étude mono-bras peut être restreinte aux patients chez lesquels le traitement a pu être mis en œuvre avec succès alors que l'étude de référence, dont le but était de décrire les pratiques, a intégré tous les patients.

Ce type de biais peut aussi prendre la forme de **biais de sélection** par exemple avec les décès précoces survenant entre l'inclusion dans l'étude et la mise en œuvre du traitement qui seront exclus que dans l'étude mono-bras et non pas dans l'étude de référence descriptive

Il est à noter qu'il est extrêmement difficile pour une étude de comparaison externe de se mettre à l'abri des biais. Cela est bien plus difficile que dans une étude comparative avec comparateur interne où, par exemple, la définition et la méthode de mesure du critère de jugement sont les mêmes entre les 2 groupes par essence. Une fois de plus, il faut rappeler qu'historiquement la méthode expérimentale au sens large, et plus particulièrement l'essai contrôlé randomisé en double aveugle, ont été construits afin d'apporter des solutions simples et efficaces à l'ensemble de ces problèmes rencontrés lors des comparaisons externes [26].

On peut donc conclure que le contrôle des biais dans une étude à comparateur externe nécessiterait de construire l'étude mono-bras en fonction de l'étude de référence : utilisation de la même définition

---

<sup>9</sup> Il convient aussi de remarquer que la problématique des biais est aussi importante que celle des différences de patients entre études mono-bras et de référence et que les méthodes d'analyse utilisée pour corriger du biais de confusion ne corrigent pas les autres biais.

du critère de jugement, de la même méthode de mesure, de la même durée de suivi, des mêmes règles d'arrêt de traitement, etc.

## 18 Le résultat suggéré par la comparaison externe doit être cliniquement pertinent

Le résultat doit être obtenu sur un critère clinique pertinent. La taille de l'effet doit elle aussi être pertinente. La balance bénéfice risque doit être appréciable et favorable qualitativement et quantitativement. La généralisabilité du résultat doit être assurée.

Comme pour toute évaluation critique de résultat d'étude clinique, la pertinence clinique est tout aussi importante que la fiabilité méthodologique du résultat. L'évaluation de la pertinence clinique d'un résultat issu d'une comparaison externe formalisée doit être évaluée avec les mêmes critères que ceux appliqués à un résultat d'essai randomisé.

## 19 Synthèses des problématiques et de leurs solutions

Problématique méthodologique spécifique (Exposant à un risque de production de résultat favorable à tort au traitement étudié)	Démonstration que doivent apporter les solutions envisagées (pour garantir la disparition du risque de conclure à tort)
Le groupe contrôle externe n'est pas un véritable contrefait documentant ce qu'aurait dû être le critère de jugement chez les mêmes patients en l'absence du traitement évalué (sous traitement de référence). C'est une problématique de biais de confusion impliquant facteurs pronostiques et modificateurs d'effet de chaque critère de jugement considéré	Les ajustements réalisés doivent donner la garantie d'une absence de confusion résiduelle. Pour cela il doit être montré que tous les facteurs pronostiques et les modificateurs de l'effet ont été pris en considération.
Malgré la réalisation d'ajustements, un biais de confusion résiduelle persiste et est à l'origine du résultat obtenu	<ul style="list-style-type: none"> <li>• L'ajustement a porté sur l'ensemble des facteurs de confusion de chacun des critères de jugement identifié par une revue systématique des facteurs pronostiques conduite suivant les standards actuels de ce type de méta-analyse<sup>10</sup> [27].</li> <li>• L'ajustement a porté sur l'ensemble des modificateurs d'effet des traitements comparés [23]</li> <li>• Et/ou une analyse quantitative de biais démontre que la taille de l'effet obtenu ne peut pas s'expliquer par le biais de confusion résiduelle</li> <li>• ET/ou l'utilisation de contrôle négatif (avec éventuellement recalibration) démontre l'absence de biais de confusion résiduel</li> </ul>
Le groupe contrôle externe a été choisi par le promoteur pour favoriser le traitement évalué (choix post hoc déterminé à partir des résultats)	<p>Choix <i>a priori</i> (avant recrutement des patients traités avec le nouveau traitement), fixé au protocole (véritable étude comparative à contrôle externe) donnant la garantie qu'il ne s'agit pas de l'étude ayant obtenu le moindre effet</p> <p>Si choix post hoc (en ayant connaissance des résultats chez les patients traités), démontrer que le choix n'est pas déterminé par les résultats en faisant une revue systématique de toutes les sources de données pouvant donner un groupe contrôle, montrer que celui utilisé était le seul possible ou montrer qu'il ne favorise pas le résultat obtenu (analyse de sensibilité avec toutes les sources possibles)</p>
Biais de mesure Extrêmement probable du fait qu'il s'agit d'une comparaison entre 2 études différentes	Aligner le protocole de l'étude à contrôle externe sur celui de l'étude servant de comparateur (même définition du critère, même méthode de mesure, etc.)

<sup>10</sup> Voir <https://methods.cochrane.org/prognosis/>

Biais de sélection	Synchroniser les débuts de suivi par rapport à l'évolution naturelle de la maladie (pour éviter un biais de temps d'immortalité) Éviter les comparaisons avant après
Biais d'attrition L'étude de comparaison est une étude descriptive où la problématique des données manquantes est moins aigue que pour les études destinées à comparer deux traitements	Utiliser un contrôle externe de qualité, ayant assuré un suivi exhaustif des patients (faible attrition)
Pertinence clinique Les études mono-bras portent souvent sur des critères intermédiaires. Les contrôles externes peuvent être trop anciens et ne pas avoir été traités avec les traitements actuellement utilisés	Le ou les critères de jugement doivent être les critères cliniques standards de la situation clinique. Le ou les traitements contrôles doivent être pertinent.
Balance bénéfice risque	La comparaison externe doit documenter la balance bénéfice risque

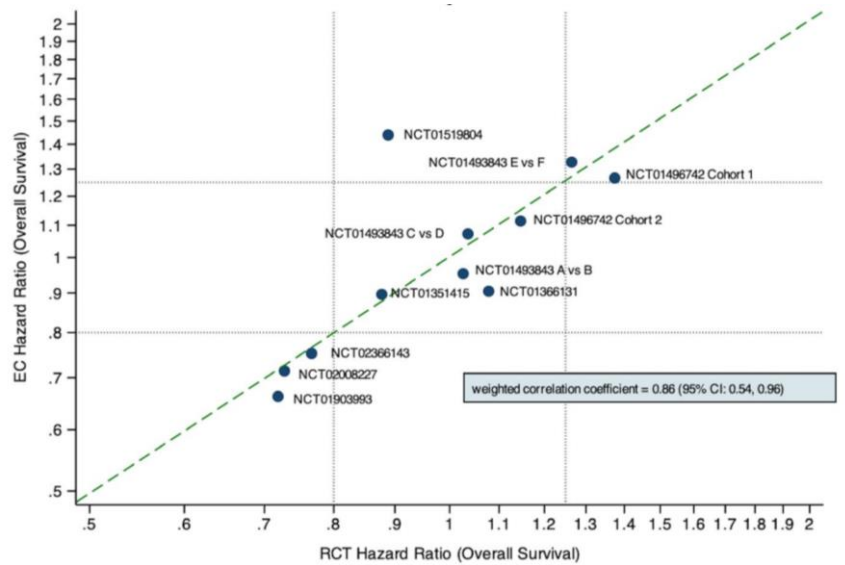
Aucun type de contrôles externes (cohorte historique, cohorte ad hoc, bras d'un essai randomisé, groupe contrôle synthétique, etc.) n'est *a priori* plus fiable. Seules comptent les réponses apportées aux problématiques méthodologiques qui entachent ces études.



## 20 Méta-recherche

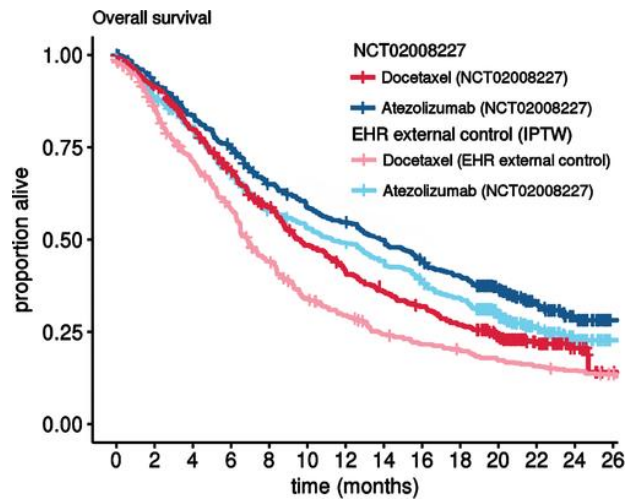
### 20.1 Fiabilité des comparaisons externes

Une étude par Carrigan et al. dans le cancer du poumon montre sur quelques cas sélectionnés une bonne aptitude des groupes contrôles synthétiques pour reproduire les résultats des essais randomisés (Figure 4) [28].



**Figure 4 – Corrélation entre les estimations produites par les groupes contrôles synthétiques (« EC Hazard Ratio ») et les essais randomisés (« RCT Hazard Ratio ») (d’après [28]).**

Comme le fait remarquer Larrouquere et al. [29], cette concordance n’apporte pas une démonstration complète de l’aptitude des groupes contrôles synthétiques à suppléer les essais randomisés, car parfois, même si les effets relatifs mesurés sont similaires, les ajustements réalisés ne reproduisent pas strictement les résultats de l’essai randomisé comme le montre la Figure 5.



**Figure 5 – Exemple des résultats obtenus par groupe contrôle synthétique comparés à ceux de l’essai randomisé (d’après [29]).**

## 20.2 Etudes monobras et enregistrement FDA

Une revue récente [30] des études monobras ayant conduit à un enregistrement accéléré par la FDA montre que ces études sont à haut risque de biais, et que les effets traitement suggérés par ces études sont de petites tailles, et potentiellement explicable par les biais.

## 21 Conclusion

L'acceptabilité d'une comparaison externe comme démonstration du bénéfice d'un nouveau traitement va dépendre de plusieurs conditions. La vérification de ces conditions va être indispensable pour conclure que les résultats ainsi produits sont suffisamment fiables, proches de ceux qui auraient été obtenus par un essai randomisé et permettent ainsi d'intégrer le nouveau traitement dans la stratégie thérapeutique.

- La comparaison externe doit être formalisée, avec l'utilisation d'une méthode de comparaisons indirecte non ancrée permettant un ajustement et débouchant sur la quantification de la taille du bénéfice
- La comparaison externe doit avoir été planifiée a priori :
  - Le jugement de l'intérêt du traitement évalué est basé sur une méthode de comparaison externe définie dans la méthode de l'étude (Il s'agit d'une véritable étude à comparateur externe et non pas d'une étude mono-bras, purement descriptive, sans objectif d'évaluation du bénéfice du traitement évalué)
- Le contrôle externe doit être justifié factuellement et non pas choisi arbitrairement :
  - La base de comparaison est une étude ou une valeur issue d'une étude (et non pas une valeur arbitraire issue d'un avis d'expert sans justification factuelle)
- Il est possible d'écarter un choix arbitraire de la référence de comparaison destiné à favoriser le traitement évalué
  - Une revue systématique de toutes les études pouvant servir potentiellement de référence est disponible
  - La référence finalement retenue n'est pas la possibilité la plus favorable au traitement évalué
  - Le choix de la base de la comparaison externe (norme ou études de référence) a été effectué indépendamment de la connaissance des résultats de l'étude mono-bras
  - Les analyses de sensibilité réalisées avec les autres études pouvant servir potentiellement de référence montrent toutes la supériorité du traitement évalué
- Les ajustements effectués permettent d'écarter un biais de confusion. L'étude donne l'assurance de l'absence d'un biais de confusion résiduel :
  - L'ajustement réalisé (MAIC ou autres) a pris en compte l'ensemble des facteurs pronostiques et des modificateurs d'effet connus pour chaque critère de jugement.
  - L'exhaustivité de la prise en compte des facteurs pronostiques est justifiée sur la base d'une revue systématique des études de facteurs pronostiques
  - L'exhaustivité de la prise en compte des modificateurs de l'effet est justifiée de manière satisfaisante (sur la base des essais comparatifs des traitements de références et de la plausibilité biologique pour le traitement évalué)
- La référence utilisée pour la comparaison externe est cliniquement loyale :
  - Cette référence correspond à des patients dont la prise en charge et les traitements reçus sont au standard actuel
- Les revues systématiques (à la recherche des groupes contrôles possible et facteurs de confusion potentiels) sont de bonne qualité
  - Les revues systématiques des potentielles études de référence et des facteurs pronostiques des critères de jugement sont satisfaisantes (recherche exhaustive, critère de sélection non arbitraire, pas d'exclusion arbitraire d'étude, évaluation du risque de biais)

- L'exposition potentielle au biais de la comparaison externe est réduite :
  - Les méthodologies des 2 études (étude monobras et étude servant de comparateur externe) sont suffisamment comparables en termes de sélection, suivi des patients et de mesure du critère de jugement pour exclure la possibilité d'une comparaison biaisée par des différences de méthode entre les études.
  - L'étude discute soigneusement tous les biais potentiels (mesure, information ou attrition) et apporte une argumentation convaincante comme quoi, quantitativement, les biais ne peuvent pas expliquer la taille de la différence observée
- Le résultat suggéré par la comparaison externe est cliniquement pertinent :
  - Le résultat est obtenu sur un critère clinique pertinent
  - la taille de l'effet est pertinente
  - La balance bénéfice risque est favorable qualitativement et quantitativement
  - la généralisabilité du résultat est assurée

## Références

- 1 Hatswell AJ, Baio G, Berlin JA, et al. Regulatory approval of pharmaceuticals without a randomised controlled study: analysis of EMA and FDA approvals 1999-2014. *BMJ open* 2016;6:e011666 doi:10.1136/bmjopen-2016-011666; PMID:27363818;
- 2 Downing NS, Aminawung JA, Shah ND, et al. Clinical trial evidence supporting FDA approval of novel therapeutic agents, 2005-2012. *JAMA* 2014;311:368-77 doi:10.1001/jama.2013.282034; PMID:24449315;
- 3 Bours MJL. A nontechnical explanation of the counterfactual definition of confounding. *J Clin Epidemiol* 2020;121:91-100 doi:10.1016/j.jclinepi.2020.01.021; PMID:32068101;
- 4 International Conference on Harmonisation of technical requirements for registration of pharmaceuticals for human use. CHOICE OF CONTROL GROUP AND RELATED ISSUES IN CLINICAL TRIALS E10.
- 5 INTERNATIONAL COUNCIL FOR HARMONISATION OF TECHNICAL REQUIREMENTS FOR PHARMACEUTICALS FOR HUMAN USE. ADDENDUM ON ESTIMANDS AND SENSITIVITY ANALYSIS IN CLINICAL TRIALS TO THE GUIDELINE ON STATISTICAL PRINCIPLES FOR CLINICAL TRIALS E9(R1).
- 6 Glassman RH, Kim G, Kahn MJ. When are results of single-arm studies dramatic? *Nat Rev Clin Oncol* 2020;17:651-52 doi:10.1038/s41571-020-00429-1; PMID:32859976;
- 7 Tang H, Foster NR, Grothey A, et al. Comparison of error rates in single-arm versus randomized phase II cancer clinical trials. *JCO* 2010;28:1936-41 doi:10.1200/JCO.2009.25.5489; PMID:20212253;
- 8 Grein J, Ohmagari N, Shin D, et al. Compassionate Use of Remdesivir for Patients with Severe Covid-19. *N Engl J Med* 2020;382:2327-36 doi:10.1056/NEJMoa2007016; PMID:32275812;
- 9 Rittberg R, Czaykowski P, Niraula S. Feasibility of Randomized Controlled Trials for Cancer Drugs Approved by the Food and Drug Administration Based on Single-Arm Studies. *JNCI Cancer Spectrum* 2021;5:pkab061 doi:10.1093/jncics/pkab061; PMID:34409254;
- 10 Kanter J, Walters MC, Krishnamurti L, et al. Biologic and Clinical Efficacy of LentiGlobin for Sickle Cell Disease. *N Engl J Med* 2022;386:617-28 doi:10.1056/NEJMoa2117175; PMID:34898139;
- 11 Tennant PWG, Arnold KF, Ellison GTH, et al. Analyses of 'change scores' do not estimate causal effects in observational data. *Int J Epidemiol* 2021 doi:10.1093/ije/dyab050; PMID:34100077;
- 12 Ozelo MC, Mahlangu J, Pasi KJ, et al. Valoctocogene Roxaparvovec Gene Therapy for Hemophilia A. *N Engl J Med* 2022;386:1013-25 doi:10.1056/NEJMoa2113708; PMID:35294811;
- 13 FDA/CDER/mccrayk. Rare Diseases: Common Issues in Drug Development: Guidance for Industry ;
- 14 International Conference on Harmonisation of technical requirements for registration of pharmaceuticals for human use. Choice of control group and related issues in clinical trials E10 2000.
- 15 Suissa S. Single-arm trials with historical controls: Study designs to avoid time-related biases. *Epidemiology* 2020 doi:10.1097/EDE.0000000000001267; PMID:33009252;
- 16 Rosenberg JE, Hoffman-Censits J, Powles T, et al. Atezolizumab in patients with locally advanced and metastatic urothelial carcinoma who have progressed following treatment with platinum-based chemotherapy: a single-arm, multicentre, phase 2 trial. *The Lancet* 2016;387:1909-20 doi:10.1016/S0140-6736(16)00561-4;
- 17 Powles T, Durán I, van der Heijden MS, et al. Atezolizumab versus chemotherapy in patients with platinum-treated locally advanced or metastatic urothelial carcinoma (IMvigor211): a multicentre, open-label, phase 3 randomised controlled trial. *The Lancet* 2018;391:748-57 doi:10.1016/S0140-6736(17)33297-X; PMID:29268948;

- 18 Signorovitch JE, Sikirica V, Erder MH, et al. Matching-adjusted indirect comparisons: a new tool for timely comparative effectiveness research. *Value Health* 2012;15:940–47 doi:10.1016/j.jval.2012.05.004; PMID:22999145;
- 19 Signorovitch JE, Wu EQ, Yu AP, et al. Comparative effectiveness without head-to-head trials: a method for matching-adjusted indirect comparisons applied to psoriasis treatment with adalimumab or etanercept. *PharmacoEconomics* 2010;28:935–45 doi:10.2165/11538370-000000000-00000; PMID:20831302;
- 20 Fox MP, Lash TL. On the Need for Quantitative Bias Analysis in the Peer-Review Process. *Am J Epidemiol* 2017;185:865–68 doi:10.1093/aje/kwx057; PMID:28430833;
- 21 Lash TL, Fox MP, Cooney D, et al. Quantitative Bias Analysis in Regulatory Settings. *Am J Public Health* 2016;106:1227–30 doi:10.2105/AJPH.2016.303199; PMID:27196652;
- 22 Lash TL, Fox MP, MacLehose RF, et al. Good practices for quantitative bias analysis. *Int J Epidemiol* 2014;43:1969–85 doi:10.1093/ije/dyu149; PMID:25080530;
- 23 David M. Phillippo, A. E. Ades, Sofia Dias, Stephen Palmer, Keith R. Abrams, Nicky J. Welton. NICE DSU technical support document 18: methods for population-adjusted indirect comparisons in submissions to NICE 2016.
- 24 Sridhara R, Mandrekar SJ, Dodd LE. Missing data and measurement variability in assessing progression-free survival endpoint in randomized clinical trials. *Clin Cancer Res* 2013;19:2613–20 doi:10.1158/1078-0432.CCR-12-2938; PMID:23669421;
- 25 Denne JS, Stone AM, Bailey-Iacona R, et al. Missing data and censoring in the analysis of progression-free survival in oncology clinical trials. *Journal of biopharmaceutical statistics* 2013;23:951–70 doi:10.1080/10543406.2013.813515; PMID:23957509;
- 26 Marks HM. *A rational therapeutics: Science and the reform of therapeutics in the United States, 1900-1990*. Cambridge England, New York: Cambridge University Press 1997 ISBN:0521581427;
- 27 Debray TPA, Damen JAAG, Snell KIE, et al. A guide to systematic review and meta-analysis of prediction model performance. *BMJ* 2017;356:i6460 doi:10.1136/bmj.i6460; PMID:28057641;
- 28 Carrigan G, Whipple S, Capra WB, et al. Using Electronic Health Records to Derive Control Arms for Early Phase Single-Arm Lung Cancer Trials: Proof-of-Concept in Randomized Controlled Trials. *Clin Pharmacol Ther* 2020;107:369–77 doi:10.1002/cpt.1586; PMID:31350853;
- 29 Larrouquere L, Giai J, Cracowski J-L, et al. Externally Controlled Trials: Are We There Yet? *Clin Pharmacol Ther* 2020;108:918–19 doi:10.1002/cpt.1881; PMID:32542679;
- 30 Ribeiro TB, Colunga-Lozano LE, Araujo APV, et al. Single-arm clinical trials that supported FDA Accelerated Approvals have modest effect sizes and at high risk of bias. *Journal of Clinical Epidemiology* 2022 doi:10.1016/j.jclinepi.2022.01.018; PMID:35093531;