



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 20 – Évaluation de la pertinence clinique

14 février 2022

Comité de rédaction et relecture (par ordre alphabétique)

Jean Luc Cracowski

Michel Cucherat

Dominique Deplanque

Behrouz Kassai

Charles Khouri

Silvy Laporte

Clara Locher

Florian Naudet

Edouard Ollier

Matthieu Roustit



[Licence Creative Commons](https://creativecommons.org/licenses/by/4.0/)

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	Pertinence du comparateur	8
2.1	Stratégie de base.....	9
2.2	Développement simultané de plusieurs molécules de la même classe.....	9
2.3	Méta-épidémiologie	9
2.4	Que faire quand le comparateur n'est pas adéquat	10
3	Pertinence clinique du critère de jugement.....	11
3.1	Interprétation du résultat obtenu sur un critère composite ?.....	12
3.1.1	Exemple d'un résultat de critère composite problématique	13
3.1.2	Exemple d'une situation ne posant pas de problème d'interprétation	13

1 Introduction

Après s'être assuré que le résultat était réel (validité interne), la lecture critique s'assure que ce résultat représente bien un réel progrès thérapeutique (une amélioration du service médical rendu si l'on reprend les termes de la commission de la transparence).

Il s'agit d'évaluer l'utilité médicale démontrée du traitement ; en quoi il amènera un changement notable, cliniquement pertinent, du devenir des patients.

En effet, un résultat peut montrer avec un degré de certitude (niveau de preuve) élevé que le traitement a un effet, mais, par exemple, sur un critère de jugement qui ne mesure pas directement le bénéfice attendu par les médecins ou les patients. Dans un cancer au stade métastatique, un résultat sur la réponse tumorale, sans effet démontré sur la survie (ou la qualité de vie) ne documente en rien si le traitement répondra aux attentes des patients en termes de survie. Dans ce cas, ce résultat n'apporte pas la preuve que le traitement est un progrès thérapeutique notable. Dans d'autres cas il pourra s'agir d'une taille d'effet trop petite, d'un comparateur non loyal ou d'une balance bénéfice risque non favorable.

Note. L'évaluation de la pertinence clinique demande de bien connaître le domaine médical de l'essai (critères de jugement pertinents, traitements déjà validés, etc.

2 Pertinence du comparateur

Le traitement comparateur (traitement du groupe contrôle) peut être un placebo ou un traitement actif.

Avec le placebo, 2 cas de figure sont possibles :

- Les patients reçoivent dans les 2 groupes un traitement de base identique. On dit que l'évaluation du nouveau traitement se fait « on top » la stratégie de base ou « en add-on ». Le but est de montrer que l'adjonction du nouveau traitement à la stratégie thérapeutique actuelle apporte un bénéfice supplémentaire aux patients.

L'essai Pegasus compare le ticagrelor à l'aspirine en prévention cardiovasculaire secondaire à distance de l'évènement initial [[10.1056/NEJMoa1500857](https://doi.org/10.1056/NEJMoa1500857)]. Tous les patients reçoivent aussi de l'aspirine dont le bénéfice a été démontré dans un essai précédent (aspirine seule versus placebo seul). La comparaison effectuée dans Pegasus est donc celle d'une bithérapie ticagrelor + aspirine versus aspirine seule (+ placebo qui est là pour assurer le double aveugle).

We randomly assigned, in a double-blind 1:1:1 fashion, 21,162 patients who had had a myocardial infarction 1 to 3 years earlier to ticagrelor at a dose of 90 mg twice daily, ticagrelor at a dose of 60 mg twice daily, or placebo. All the patients were to receive low-dose aspirin.

- Les patients du groupe contrôle ne reçoivent que le placebo. L'évaluation ne se fait pas par-dessus d'autres traitements, car aucun traitement n'a démontré son intérêt dans cette situation. Le placebo est alors un comparateur pertinent, car il n'existe pas de traitement de référence. En revanche, si un traitement a déjà montré son intérêt, faire un essai versus placebo n'est pas pertinent, car 1) cela représente une perte de chance pour les patients du groupe contrôle (sauf si l'abstention thérapeutique est aussi une option) et 2) ne répond pas à la question clinique qui se pose dans cette situation : peut-on avoir un traitement plus efficace à celui déjà disponible ?

Several studies have tested the administration of systemic therapy with curative intent after patients had disease control with chemoradiotherapy. However, to date, these therapies have proved ineffective. PACIFIC was a randomized, placebo-controlled, phase 3 trial evaluating the immune checkpoint inhibitor durvalumab [[10.1056/NEJMoa1809697](https://doi.org/10.1056/NEJMoa1809697)]

Lorsque le comparateur est un traitement actif, l'essai répond à la question : le remplacement du traitement de référence par le nouveau traitement apporte-t-il un bénéfice supplémentaire aux patients.

Warfarin and other vitamin K antagonists are highly effective treatments, reducing the risk of stroke by about two thirds, but their use is limited by a narrow therapeutic range, drug and food interactions, required monitoring, and risk of bleeding. In the ARISTOTLE trial, we compared apixaban with warfarin for the prevention of stroke or systemic embolism in patients with atrial fibrillation and at least one additional risk factor for stroke. [[10.1056/NEJMoa1107039](https://doi.org/10.1056/NEJMoa1107039)]

Le traitement comparateur actif doit être le meilleur traitement disponible, utilisé de manière optimale. Autrement la comparaison est déloyale, car ne s'effectuant pas par rapport à la meilleure référence possible.

Note : Cette évaluation demande des connaissances sur le domaine médical, et au niveau du 2^{ème} cycle elle ne sera abordée qu'avec la mise à disposition d'une documentation permettant de juger de la situation sans connaissance préalable.

2.1 Stratégie de base

L'évaluation de la pertinence clinique du traitement comparateur ne s'arrête pas stricto-sensu au traitement contrôle mais porte sur l'ensemble des traitements que reçoivent les patients du groupe contrôle. La plupart des essais évaluent en fait l'ajout d'un nouveau traitement à la stratégie thérapeutique de base versus la stratégie de base seule. La question de la pertinence clinique du traitement du groupe contrôle devient alors la pertinence de la stratégie de base. Correspond-elle à la meilleure stratégie validée ?

La crainte à ce niveau est que l'ajout d'un nouveau produit montre un éventuel intérêt seulement parce que les patients ne reçoivent pas à la base un traitement optimum. Dans ce cas il « reste de la place » pour montrer le bénéfice de l'ajout. Avec un traitement de base optimum, la totalité du bénéfice qu'il est possible d'apporter aux patients est déjà donné et l'ajout d'un nouveau traitement ne sert à rien. Avec un traitement de base optimum, un essai d'ajout du nouveau traitement serait négatif.

2.2 Développement simultané de plusieurs molécules de la même classe

Il existe une certaine iniquité lorsque plusieurs molécules du même mécanisme d'action se développent en parallèle. La première molécule n'aura qu'à montrer sa supériorité au placebo (ou à un traitement actif moins efficace) tandis qu'en toute rigueur, les molécules suivantes devraient montrer leur supériorité à cette première molécule.

En pratique, il est admis que

Par exemple, les essais des trois premiers anticoagulants oraux directs dans la fibrillation auriculaire (dabigatran, rivaroxaban et apixaban) ont tous utilisé la warfarine avec un INR entre 2 et 3 comme traitement contrôle. ???

2.3 Méta-épidémiologie

En oncologie par exemple, il est de plus en plus fréquent que l'optimalité du traitement du groupe contrôle soit discuté dans le sens d'une remise en cause [1, 2].

Une étude de méta-épidémiologie portant sur les 95 essais randomisés de phase 3 ayant conduit à un enregistrement de traitements anti-cancer par la FDA entre 2013 et 31 juillet 2018 que 16 (17%) d'entre eux avaient utilisé un traitement contrôle non optimal [2] :

4 (25%) trials omitted active treatment in control arm by limiting investigator's choice, 11 (63%) trials omitted active treatment in the control arm by using a control agent known to be inferior to other available agents or not allowing combinations, and 1 (13%) trial used a previously used treatment in the control arm with a known lack of benefit associated with reexposure.

Des résultats similaires ont été relevés dans les essais récents du myélome multiple [3].

2.4 Que faire quand le comparateur n'est pas adéquat

Peu de solutions vraiment satisfaisantes sont disponibles lorsque le comparateur n'est pas le traitement de référence. La seule piste possible est celle des comparaisons indirectes par méta-analyse en réseau par exemple. L'idée serait de produire le résultat de la comparaison du nouveau traitement avec le traitement de référence.

Cette solution est d'autant plus envisageable que l'inadéquation du comparateur est due aux évolutions thérapeutiques durant la réalisation de l'étude et n'est pas un choix délibéré.

Si le nouveau traitement N a été évalué versus le traitement de référence C alors que durant la même période un autre traitement N' est devenu le nouveau traitement de référence en montrant sa supériorité au même comparateur C, le but de la comparaison indirecte sera d'extrapoler les résultats de N vers N' en utilisant les essais N vs C et N' vs C (cf. chapitre ???)

3 Pertinence clinique du critère de jugement

La démonstration de l'intérêt clinique d'un traitement nécessite l'utilisation d'un critère clinique.

La mise en évidence d'un effet sur un critère intermédiaire est nettement insuffisante pour justifier l'utilisation d'un traitement en pratique

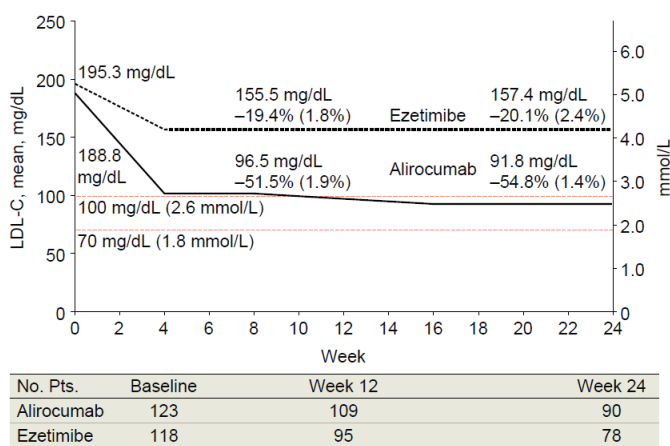
Les critères cliniques correspondent directement à la problématique de la pathologie à traiter. Il s'agit souvent d'évènements cliniques (décès, AVC), mais aussi de signes fonctionnels (douleur, succès thérapeutique sur une échelle de dépression, etc.).

Les critères intermédiaires sont proches du mécanisme d'action des traitements et correspondent le plus souvent à des paramètres biologiques (ou d'imagerie). Mais un effet du traitement sur un critère intermédiaire ne s'accompagne pas toujours d'un bénéfice sur le critère clinique (cf. section **Erreur ! Source du renvoi introuvable.**). Pour cette raison, les critères intermédiaires sont insuffisants pour évaluer l'intérêt clinique des traitements.

Tableau 1 – Exemples de critères cliniques et intermédiaires

Pathologie	Critères cliniques	Critères intermédiaires
Ostéoporose	Fractures vertébrales Fractures os long	Densité osseuse
Hypertension	Évènements cardiovasculaires (infarctus, AVC, décès de cause cardiovasculaires)	Pression artérielle
COVID-19	Mortalité, aggravation clinique, décès ou ventilation mécanique	Charge virale
Arthrose	Échelle de douleur (EVA), Indice algo-fonctionnel de Lequesne	Pincement artriculaire

L'alirocumab est un hypolipémiant de dernière génération. Les premiers essais réalisés avec cette molécule avaient comme critère de jugement le LDL-cholesterol, comme, par exemple dans les essais Odyssey ALTERNATIVE [[10.1016/j.jacl.2015.08.006](https://doi.org/10.1016/j.jacl.2015.08.006)] ou ODYSSEY COMBO II [[10.1111/dom.12909](https://doi.org/10.1111/dom.12909)].



Mais ces essais sur critères intermédiaires n'étaient pas suffisamment cliniquement pertinents pour apporter la preuve de l'intérêt du traitement. Un grand essai de morbi-mortalité, ODYSSEY OUTCOMES [[10.1056/NEJMoa1801174](https://doi.org/10.1056/NEJMoa1801174)], incluant 18,924 avec un suivi médian de 2.8 ans, a donc été entrepris. L'introduction de la publication justifie ainsi la réalisation de cet essai :

“Studies have shown that mutations conveying gain or loss of function of PCSK9 result in a higher or lower level of LDL cholesterol, respectively, which in turn is associated with a corresponding higher or lower risk of incident coronary heart disease. These findings have led to the development of monoclonal antibodies to PCSK9 that produce substantial reductions in LDL cholesterol when administered alone or with a statin. To date, the potential for a PCSK9 antibody to reduce cardiovascular risk after an acute coronary syndrome remains undetermined. In the ODYSSEY OUTCOMES trial, we tested the hypothesis that treatment with alirocumab, a fully human monoclonal antibody to PCSK9, would result in a lower risk of recurrent ischemic cardiovascular events than placebo ... “.

3.1 Interprétation du résultat obtenu sur un critère composite ?

Un critère composite est un critère qui regroupe plusieurs entités cliniques (plusieurs types d'évènements cliniques) comme les évènements cardiovasculaires dont les composantes sont les infarctus du myocarde non mortels, les AVC non mortels et les décès de causes cardiovasculaires.

Un patient sera comptabilisé comme « ayant fait » le critère composite à partir du moment où il présente un des évènements composants. Ainsi, dans le tableau de résultat de l'essai, le nombre de critères composites est le nombre de patients ayant présentés au moins un des évènements composants.

End Point	Ticagrelor, 90 mg (N = 7050)	Ticagrelor, 60 mg (N = 7045)	Placebo (N = 7067)
	<i>number (percent)</i>		
Cardiovascular death, myocardial infarction, or stroke	493 (7.85)	487 (7.77)	578 (9.04)
Death from coronary heart disease, myocardial infarction, or stroke	438 (6.99)	445 (7.09)	535 (8.33)
Cardiovascular death or myocardial infarction	424 (6.79)	422 (6.77)	497 (7.81)
Death from coronary heart disease or myocardial infarction	350 (5.59)	360 (5.75)	429 (6.68)
Cardiovascular death	182 (2.94)	174 (2.86)	210 (3.39)
Death from coronary heart disease	97 (1.53)	106 (1.72)	132 (2.08)
Myocardial infarction	275 (4.40)	285 (4.53)	338 (5.25)
Stroke			
Any	100 (1.61)	91 (1.47)	122 (1.94)
Ischemic	88 (1.41)	78 (1.28)	103 (1.65)

Figure 1 – Exemple de tableau de résultat rapportant un critère composite

La première ligne présente les résultats du critère évènement cardiovasculaires. Dans le groupe ticagrelor 90mg, 493 patients ont présenté au moins un évènement composant ce critère : un décès d'ordre cardiovasculaire, un AVC non mortel ou un infarctus non mortel. La présence de cooccurrences (patients ayant présenté plusieurs de ces évènements, par exemple un premier infarctus puis 6 mois après un second mortel) est attestée par le fait de la somme des évènements composants ne donne pas le nombre du composite (100 AVC + 275 infarctus + 182 décès CV =557 pour 493 patients pour le composite) [[10.1056/NEJMoa1500857](https://doi.org/10.1056/NEJMoa1500857)]

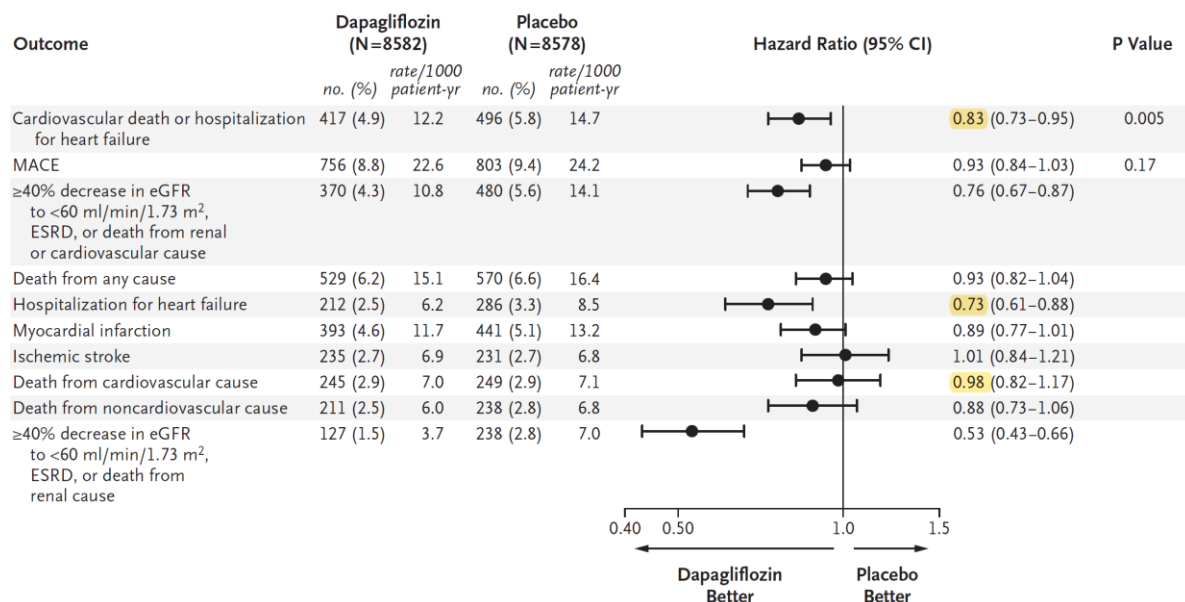
Une limite des critères composites est qu'ils peuvent regrouper des évènements dont la signification clinique est très inégale (en termes de gravité par exemple). Ces composantes les moins cliniquement

signifiantes peuvent être majoritaires, ce qui fait que l'effet observé au niveau du critère composite représente principalement l'effet du traitement sur l'évènement le moins cliniquement significatif.

Les critères composites regroupent fréquemment des évènements mortels et non mortels. Souvent ce qui fait la valeur d'un nouveau traitement dans une aire thérapeutique déjà pourvue est une réduction de la mortalité. D'où l'enjeu de pouvoir conclure d'une façon ou d'une autre à cette réduction. Quand un critère de morbi-mortalité est réduit par le traitement, la conclusion que le traitement a réduit la fréquence du critère « décès ou évènement non mortels » peut être facilement compris comme une réduction « des décès » et « des évènements mortels ». Cependant, dans certains cas, cette réduction du composite n'a été obtenue que par un effet sur les évènements non mortels. L'impression que donne alors la conclusion est excessive par rapport à la réalité des résultats.

3.1.1 Exemple d'un résultat de critère composite problématique

Dans un essai [[10.1056/NEJMoa1812389](#)] chez le patient diabétique de type 2, le critère de jugement principal était défini par « *The primary efficacy outcomes were MACE and a composite of cardiovascular death or hospitalization for heart failure.* ». Les résultats obtenus sont les suivants :



La réduction (HR=0.83) du critère composite décès cardiovasculaire ou hospitalisation a été obtenu uniquement par une réduction des hospitalisations (HR=0.73). Aucune tendance n'est observée sur les décès (HR=0.98). Finalement le résultat sur le composite ne représente qu'un effet sur les hospitalisations et non pas un effet homogène sur les 2 composantes (cf. ci-dessous). Conclure à une réduction des décès ou des hospitalisations devient excessif contrairement à une situation où les 2 composantes auraient été réduites de la même manière.

Pour éviter cette confusion, l'article conclut de la façon suivante : « *In patients with type 2 diabetes who had or were at risk for atherosclerotic cardiovascular disease, treatment with dapagliflozin ... did result in a lower rate of cardiovascular death or hospitalization for heart failure, a finding that reflects a lower rate of hospitalization for heart failure* ».

3.1.2 Exemple d'une situation ne posant pas de problème d'interprétation

Dans un essai chez des patients insuffisants cardiaques [[10.1056/NEJMoa1911303](#)], le critère de jugement principal était : « *The primary outcome was a composite of worsening heart failure*

(hospitalization or an urgent visit resulting in intravenous therapy for heart failure) or cardiovascular death. ».

Une réduction du critère composite a été obtenue identique en taille (HR=0.74) avec la réduction des hospitalisations (HR=0.70) et celle des décès (HR=0.82). La conclusion que le traitement réduit le critère décès ou hospitalisation n'est donc pas excessif. Les résultats sont en concordance avec le sens trivial de la notion de composite.

Table 2. Primary and Secondary Cardiovascular Outcomes and Adverse Events of Special Interest.*

Variable	Dapagliflozin (N = 2373)		Placebo (N = 2371)		Hazard or Rate Ratio or Difference (95% CI)	P Value
	values	events/100 patient-yr	values	events/100 patient-yr		
Efficacy outcomes						
Primary composite outcome — no. (%) [†]	386 (16.3)	11.6	502 (21.2)	15.6	0.74 (0.65 to 0.85)	<0.001
Hospitalization or an urgent visit for heart failure	237 (10.0)	7.1	326 (13.7)	10.1	0.70 (0.59 to 0.83)	NA
Hospitalization for heart failure	231 (9.7)	6.9	318 (13.4)	9.8	0.70 (0.59 to 0.83)	NA
Urgent heart-failure visit	10 (0.4)	0.3	23 (1.0)	0.7	0.43 (0.20 to 0.90)	NA
Cardiovascular death	227 (9.6)	6.5	273 (11.5)	7.9	0.82 (0.69 to 0.98)	NA

La conclusion de l'article est "Among patients with heart failure and a reduced ejection fraction, the risk of worsening heart failure or death from cardiovascular causes was lower among those who received dapagliflozin than among those who received placebo".

La démonstration d'un bénéfice sur un critère composite ne nécessite pas que les composantes soient statistiquement significatives individuellement. Cela montre bien qu'une réduction sur un critère « de morbi ou de mortalité » n'a pas valeur d'une démonstration sur la morbidité et d'une démonstration sur la mortalité. Si dans un domaine particulier, il est nécessaire d'avoir une démonstration spécifique d'un bénéfice en termes de mortalité par exemple (car d'autres produits ont déjà montré leur aptitude à réduire la mortalité), une démonstration ne portant que sur le composite ne sera pas suffisante pour apporter la preuve de l'intérêt du traitement. Il faudra, dans ce cas, une démonstration spécifique de la réduction de la mortalité. Les décès devront alors être inclus dans le plan de contrôle du risque alpha global (hiérarchisation le plus souvent).

Si le critère composite a pour objectif d'évaluer le bénéfice net (cf. section **Erreur ! Source du renvoi introuvable.**) en regroupant des événements que le traitement prévient et des effets indésirables qu'il induit, l'homogénéité des effets n'est plus une condition d'acceptabilité du résultat, vu que par construction des effets opposés sont attendus.

Dans un essai d'un anticoagulant dans la FA [[10.1056/NEJMoa1107039](#)], un critère de bénéfice net a été utilisé regroupant les AVC et embolies systémiques que le traitement cherche à prévenir avec les hémorragies majeures. Une réduction de ce critère est obtenue montrant que les hémorragies ne contrebalancent pas le bénéfice obtenu. La balance bénéfice risque du traitement est favorable.

Outcome	Apixaban Group (N=9088)		Warfarin Group (N=9052)		Hazard Ratio (95% CI)	P Value
	Patients with Event	Event Rate	Patients with Event	Event Rate		
	<i>no.</i>	<i>%/yr</i>	<i>no.</i>	<i>%/yr</i>		
Net clinical outcomes						
Stroke, systemic embolism, or major bleeding	521	3.17	666	4.11	0.77 (0.69–0.86)	<0.001
Stroke, systemic embolism, major bleeding, or death from any cause	1009	6.13	1168	7.20	0.85 (0.78–0.92)	<0.001

Références

- 1 Hilal T, Gonzalez-Velez M, Prasad V. Limitations in Clinical Trials Leading to Anticancer Drug Approvals by the US Food and Drug Administration. *JAMA Intern Med* 2020;180:1108–15
doi:10.1001/jamainternmed.2020.2250; PMID:32539071;
- 2 Hilal T, Sonbol MB, Prasad V. Analysis of Control Arm Quality in Randomized Clinical Trials Leading to Anticancer Drug Approval by the US Food and Drug Administration. *JAMA Oncol* 2019;5:887–92
doi:10.1001/jamaoncol.2019.0167; PMID:31046071;
- 3 Mohyuddin GR, Koehn K, Sborov D, et al. Quality of control groups in randomised trials of multiple myeloma enrolling in the USA: a systematic review. *The Lancet Haematology* 2021;8:e299-e304
doi:10.1016/S2352-3026(21)00024-7;