

# 10. Les essais randomisés en grappe

## *Un exemple en santé maternelle et infantile*

ALEXANDRE DUMONT

---

### Les essais randomisés en grappe

---

#### Définition de la méthode

Les essais randomisés en grappe permettent d'évaluer l'impact des interventions en santé mise en œuvre à l'échelle d'un groupe d'individus, tels que les patientes qui accouchent dans un même hôpital ou les habitants d'un même village.

#### Forces de la méthode

- permet d'éviter le biais de contamination par l'intervention des individus (ou cibles) de deux grappes différentes
- permet d'avoir les effets de l'intervention avec un niveau de preuve élevé
- les essais contrôlés en grappe sont le plus souvent des études expérimentales pragmatiques qui évaluent les interventions dans la vie réelle

#### Défis de la méthode

- l'intervention doit être la plus standardisée possible
- défis statistiques liés au fait que l'unité de randomisation et d'analyse n'est pas la même
- défis dans la sélection des grappes qui doivent être représentatives du système de santé où se déroule la recherche

---

La mise en œuvre des politiques de santé nécessite d'expérimenter régulièrement de nouvelles interventions adaptées aux contextes et censées faciliter l'implantation de bonnes pratiques ayant un fort impact sur la santé des populations. Dans la lutte contre la mortalité maternelle, la plupart des politiques de santé périnatale dans le monde sont basées sur l'accouchement assisté par du personnel qualifié et le traitement des



*Maternité en RDC. Crédit « Action santé des femmes ».*

complications obstétricales lorsqu'elles surviennent pendant le travail ou l'accouchement. Ces traitements sont connus, mais le défi pour la plupart des systèmes de santé en développement est de les mettre en œuvre en temps utile. De nombreuses initiatives ont été testées dans les pays à faibles ou moyens revenus (PFMR) pour mobiliser les communautés ou améliorer la filière de soins et les pratiques médicales afin que les femmes puissent accéder plus rapidement à des services obstétricaux de qualité.

Quelques évaluations expérimentales ou quasi-expérimentales ont permis de fournir des premières preuves sur l'efficacité et la mise en œuvre des interventions qui visent à réduire la mortalité maternelle. Mais ces informations sont encore fragmentaires et peu d'études d'impact de bonne qualité ont été réalisées dans les PFMR en comparaison avec les pays où les niveaux de ressources sont plus élevés. Il est pourtant nécessaire de produire et de diffuser plus de données probantes sur l'efficacité de ces interventions, qu'il s'agisse de programmes nationaux ou d'initiatives plus localisées. Ceci doit permettre aux décideuses et décideurs et aux responsables de programmes de santé maternelle de concevoir des politiques publiques plus efficaces.

Le niveau de preuve des études d'impact en population peut être comparé à celui des études cliniques pour les traitements médicamenteux (voir Tableau 1). Les études avec le niveau de preuve le plus élevé sont les essais contrôlés randomisés (ECR) dans lesquels la plupart des biais de sélection des participant-e-s sont contrôlés, permettant ainsi d'affirmer que les changements observés sont bien attribuables à l'intervention et non aux différences entre les individus ou les pratiques de soins. Il existe deux catégories d'ECR : les ECR individuels où les sujets sont tirés au sort individuellement pour participer à différents groupes d'intervention (randomisation individuelle) et les ECR en grappes dans lesquels on randomise des groupes de sujets (randomisation en grappe).

**Tableau 1. Niveaux de preuve scientifique et grades des recommandations (Source : Hautes Autorités de Santé, France)**

<b>Niveau de preuve scientifique fourni par la littérature (études thérapeutiques)</b>	<b>Grade des recommandations</b>
Niveau 1 (NP1) Essais comparatifs randomisés de forte puissance. Méta-analyse d'essais comparatifs randomisés Analyse de décision basée sur des études bien menées	Preuve scientifique établie A
Niveau 2 (NP2) Essais comparatifs randomisés de faible puissance. Études comparatives non randomisées bien menées Études de cohorte	Présomption scientifique B
Niveau 3 (NP3) Études cas-témoins	Faible niveau de preuve C

Les ECR individuels sont peu fréquents dans les PFMR pour trois raisons essentielles : (i) le coût de ces études est élevé, car elles nécessitent le plus souvent le recrutement de patient-e-s dans de multiples centres pendant plusieurs années; (ii) la plupart des interventions qu'il faudrait tester dans un contexte à faibles ressources ont déjà une efficacité connue dans les pays à ressources élevées; (iii) la randomisation de sujets n'est parfois pas possible pour évaluer des interventions dans la communauté pour des raisons logistiques, administratives, voire éthiques. Un autre inconvénient des ECR individuels est lié au fait que le recrutement des sujets est souvent très sélectif en raison des multiples critères d'inclusion et d'exclusion et les conditions dans lesquelles les interventions sont mises en œuvre sont trop standardisées. Les résultats obtenus par un ECR sont alors difficilement généralisables à d'autres contextes que celui dans lequel l'étude a été menée.

L'approche utilisée dans les essais randomisés en grappes est beaucoup plus pragmatique que l'ECR individuel (Tableau 2). L'intervention testée ne cible pas directement l'individu, mais une unité « sociale » indépendante (la famille, le village, le médecin, l'hôpital...) pour faire changer les comportements ou introduire de nouvelles pratiques dans la « vie réelle ».

Tableau 2. Essais randomisés individuels et en grappes

	ECR individuel	ECR en grappes
<b>Unité d'intervention et de randomisation</b>	Sujet	Groupe de sujets
<b>Type d'intervention</b>	Traitement	Programme d'intervention
<b>Objectif</b>	Améliorer l'état de santé	Changer les comportements
<b>Risque de contamination</b>	Élevé	Faible
<b>Contexte d'étude</b>	Très standardisé Milieu clinique	« Vie réelle » En population
<b>Coût</b>	Élevé	Élevé

## Défis des essais contrôlés randomisés en grappes

Si les essais randomisés en grappes offrent plusieurs avantages dans les PFMR par rapport aux essais individuels, ils présentent quelques défis méthodologiques qui tiennent au fait que l'unité d'analyse est différente de l'unité de randomisation ou d'intervention (Figure 1).

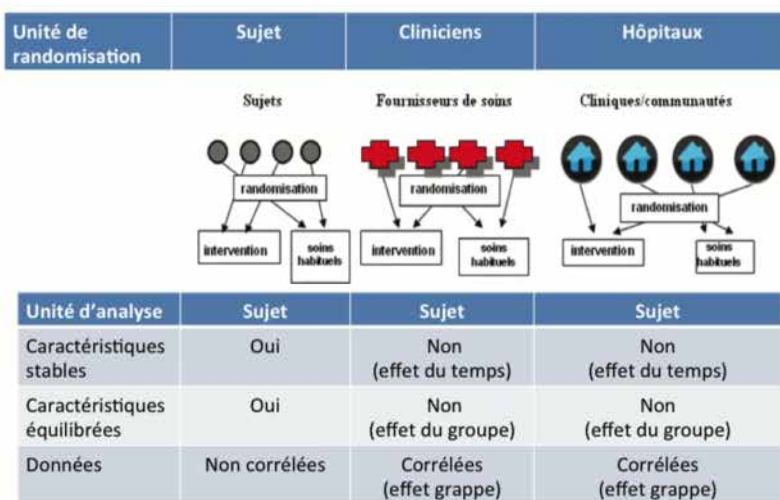


Figure 1. Implications des essais randomisés contrôlés en grappes

En randomisant des grappes, on se met en situation de recueillir des informations au niveau des individus dont les caractéristiques ne sont pas nécessairement stables au cours du temps, ni équilibrées entre les bras de l'essai. De plus, ces données sont corrélées au sein de chaque grappe. Cette situation conduit à trois conséquences importantes à considérer pour l'analyse des résultats d'un essai en grappe.

Premièrement, les caractéristiques des individus au sein des grappes, mais aussi la performance globale du système de santé, peuvent évoluer pendant la période d'étude. C'est ce qui explique les changements observés des indicateurs de santé dans les grappes où aucune intervention extérieure n'est mise en œuvre (groupe contrôle) et que nous nommerons « l'effet du temps ». Par exemple, le ratio de mortalité maternelle a chuté de près de 44 % au cours des 25 dernières années à l'échelle mondiale. De nombreux pays ont mis en place des politiques de santé qui ont contribué à cette évolution favorable. Ne pas prendre en compte la tendance naturelle des indicateurs de santé maternelle dans une étude d'impact d'une nouvelle intervention reviendrait à attribuer les changements observés uniquement à cette intervention et non à l'effet combiné du temps et de l'intervention. C'est pourquoi il est indispensable de comparer les résultats observés dans un groupe expérimental à ceux observés dans un groupe contrôle sans intervention extérieure. Nous verrons comment l'approche « différence des différences » permet dans ce cas d'individualiser l'effet de l'intervention de l'effet du temps.

Une autre implication méthodologique de la randomisation en grappe est en lien avec un déséquilibre fréquent des caractéristiques des individus entre les groupes de comparaison. En effet, la randomisation par grappe conduit de façon quasi systématique à une situation où la randomisation est antérieure à l'inclusion des sujets. Les caractéristiques des grappes sont alors réparties de façon homogène entre les groupes si la randomisation est effectuée dans les règles de l'art. Ce n'est pas nécessairement le cas en ce qui concerne les caractéristiques des individus qui constituent chaque grappe. Le déséquilibre qui en résulte, nommé ici « l'effet du groupe », peut expliquer les différences des indicateurs de santé entre les groupes de comparaison avant même d'avoir démarré l'intervention. Le risque de ne pas prendre en compte ce déséquilibre serait d'attribuer l'effet de l'intervention à la différence initiale entre les groupes, alors que l'intervention n'a par ailleurs aucune efficacité.

Enfin, en randomisant des grappes, on se met en situation de recueillir des données corrélées. Les sujets qui composent une unité « sociale » ou grappe ne peuvent être considérés comme indépendants les uns des autres. En effet les résultats observés chez deux sujets d'une même grappe auront tendance à être plus « similaires » que si les résultats étaient associés à deux sujets provenant de deux grappes distinctes (donc indépendantes). Si l'on ne tenait pas compte de la corrélation induite par le schéma expérimental (l'effet grappe), la puissance des tests statistiques se trouverait sous-estimée, les étendues des intervalles de confiance faussement diminuées et les conclusions de l'étude erronées.

Dans le cas de l'essai QUARITE réalisé au Sénégal et au Mali, nous avons randomisé des hôpitaux pour intervenir sur les professionnels de santé en faisant l'hypothèse que l'amélioration de la qualité des soins allait entraîner une diminution de la mortalité maternelle. Nous verrons, à travers cet exemple, comment l'utilisation des méthodes statistiques appropriées permet d'analyser correctement ces données.

### **Contexte et programme évalué : l'essai QUARITE pour réduire la mortalité maternelle hospitalière**

Malgré la baisse de la mortalité maternelle dans le monde, celle-ci reste toujours très élevée dans de nombreux pays à faibles ou moyens revenus. Une large part des décès maternels est due à un traitement trop tardif ou inapproprié des complications obstétricales. Il est donc urgent de mettre en œuvre des interventions efficaces pour remédier à cette situation. Ces interventions consistent en général à mobiliser les communautés pour améliorer le recours aux soins en temps utile et à améliorer la filière de soins et les pratiques médicales aux différents niveaux de la pyramide sanitaire.

Nous avons testé par un ECR en grappes (essai QUARITE), entre 2007 et 2011, une approche combinant les revues de cas de mortalité maternelle dans les établissements de santé (ou audit de décès) et la formation continue du personnel de santé (Dumont *et al.* 2013). L'unité de randomisation et d'intervention était l'hôpital de référence dans lequel les complications obstétricales les plus sévères étaient prises en charge, au Mali comme au Sénégal. L'intervention ciblait les professionnel-le-s de santé qui travaillaient dans le même service (maternité). L'intervention d'une durée de deux ans

consistait à former initialement des leaders d'opinion locaux aux pratiques optimales et aux audits de décès maternels (formation de formateurs et formatrices). Les leaders d'opinion formaient ensuite leurs équipes (médecins, sages-femmes et infirmières qualifiées) dans chaque hôpital et implantaient les audits de décès maternels. Un facilitateur ou une facilitatrice externe visitait tous les trois mois chaque centre pour vérifier si l'agenda de formation était bien respecté et superviser une séance d'audit. L'audit des décès maternels avait pour objectif de rechercher de manière qualitative et approfondie les causes et les circonstances entourant les décès maternels survenus dans les établissements de santé. Dans un premier temps, ces décès étaient recensés et analysés dans les établissements. Mais lorsque cela était possible, ces études s'attachaient également à identifier les facteurs intervenant en dehors de l'hôpital, dans le système de santé ou dans la communauté, et ayant contribué au décès. La formation du personnel de santé ciblait les problèmes de prise en charge qui avaient été identifiés lors des séances d'audit afin de contribuer à l'amélioration de la qualité des soins dans les hôpitaux participants.

## Protocole d'évaluation

Pour mesurer les effets de l'intervention QUARITE sur la mortalité maternelle dans les 46 hôpitaux de référence participant à l'essai, nous avons enregistré toutes les naissances dans ces centres pendant quatre ans, entre octobre 2007 et octobre 2011. L'identification des morts maternelles a été réalisée à partir de multiples sources d'information comme les registres d'accouchement, du bloc opératoire, d'hospitalisation ou de la morgue et les dossiers cliniques disponibles. Le critère de jugement principal de l'essai était le taux de mortalité maternelle estimé par le nombre de décès maternels enregistré dans l'hôpital divisé par le nombre d'accouchements pendant la même période.

La collecte des données pendant quatre ans a permis d'estimer cet indicateur avant la mise en œuvre de l'intervention (année 1), pendant la période d'intervention (année 2 et 3) et après la fin de la l'intervention (année 4).

Les hôpitaux ont été randomisés à la fin de la première année de collecte, juste avant le démarrage de l'intervention, entre un groupe

expérimental exposé au programme et un groupe contrôle sans intervention extérieure. La figure 2 présente l'évolution de la mortalité maternelle dans les deux groupes pendant la période d'étude. Elle a diminué de 10,4 décès pour 1 000 accouchements à 6,8 décès pour 1000 dans le groupe expérimental entre les périodes « avant » et la période « après » l'intervention. Cette diminution a été moins marquée dans le groupe contrôle : entre 8,1 et 7,1 pour 1 000 accouchements.

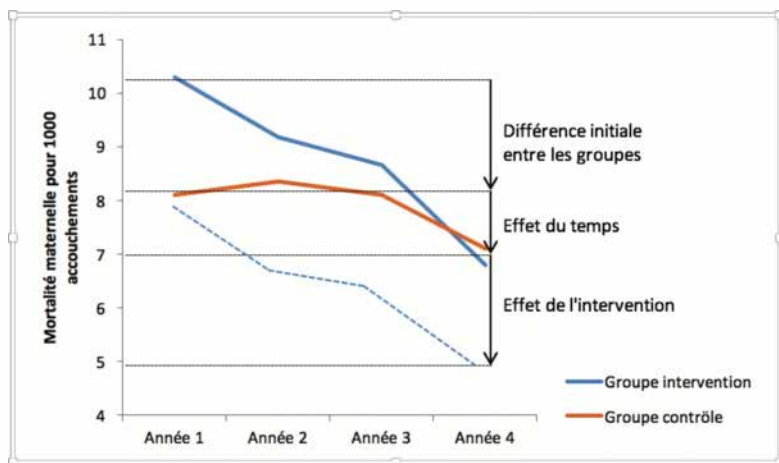


Figure 2. Évolution de la mortalité maternelle dans les deux groupes – essai QUARITE  
 Note : la ligne pointillée représente l'évolution de la mortalité maternelle dans le groupe d'intervention si son niveau de base (avant intervention) eut été équivalent à celui du groupe contrôle

## Application de la méthode

Prise en compte de l'effet du temps : l'approche « différence des différences »

La différence de mortalité maternelle entre les deux périodes dans le groupe contrôle représente la tendance naturelle de la mortalité maternelle dans les hôpitaux de référence du Mali et du Sénégal (Figure 2 ci-dessus).



Cette évolution est peu marquée (-1,0 Décès pour 1 000 accouchements) sur la période d'étude, mais elle n'est pas négligeable. Elle témoigne de l'effet du changement des caractéristiques des patientes au cours du temps, mais aussi des changements intervenus dans le système de santé. La gratuité de la césarienne initiée en 2005 au Sénégal puis en 2006 au Mali a probablement facilité l'accès aux services de santé et amélioré progressivement l'issue des accouchements entre 2007 et 2011. De plus, la supervision régulière des équipes de soins pendant la période de l'essai pour renforcer la qualité des données collectées a peut-être contribué à améliorer les soins courants dans les hôpitaux des deux groupes et la prise en charge des patientes. Pour toutes ces raisons, les résultats de santé maternelle se sont améliorés au cours de l'étude dans le groupe contrôle alors que les hôpitaux de ce bras n'ont reçu aucune intervention de notre part. La méthode « différence des différences » (DD) permet de prendre en compte cet effet du temps dans l'étude d'impact.

Dans l'approche DD, l'effet de l'intervention est mesuré par la différence des changements observés entre les deux groupes (voir Tableau 3 ci-dessous). Les résultats peuvent être exprimés en termes de différence des risques de mortalité ou de rapport de cotes (RC).

En termes de différence des risques, l'effet de l'intervention correspond à la diminution des taux de mortalité maternelle dans le groupe expérimental (- 3,5 décès pour 1000) moins la diminution de mortalité dans le groupe contrôle (- 1,0 pour 1000). Dans cet exemple, la DD est donc estimée à  $(-3,5) - (-1,0) = - 2,5$  décès pour 1 000 accouchements (IC à 95 % = - 3,5 ; - 1,5 pour 1 000 accouchements). En d'autres termes, l'intervention dans un hôpital réalisant 1 000 accouchements permettrait d'éviter 2,5 décès maternels par rapport à un autre hôpital de même volume d'activité, mais ne bénéficiant pas du programme. Cet effet est statistiquement significatif, car l'intervalle de confiance à 95 % de la DD ne contient pas 0.

Lorsque les résultats sont exprimés en rapport de cotes (RC), l'effet de l'intervention correspond au ratio entre les rapports de cotes des deux groupes. Dans notre cas, la cote pour un groupe donné à une période donnée est le nombre de décès maternels divisé par le nombre de femmes vivantes avant leur sortie de l'hôpital. Le rapport de cotes (RC) entre les deux périodes correspond donc à la variation du risque de mortalité maternelle entre l'année 1 et l'année 4. La mortalité ayant diminué dans les deux groupes entre les deux périodes d'étude, les RC sont inférieurs à 1 (Tableau 2). Toutefois,

cette diminution a été plus importante dans le groupe expérimental (RC = 0,66) que dans le groupe contrôle (RC = 0,89). Le rapport entre les deux RC, ici égal à 0,73 (IC à 95 % = 0,61 ; 0,91), mesure l'effet de l'intervention qui, là aussi, est statistiquement significatif, car l'intervalle de confiance du ratio des RC ne contient pas 1. En d'autres termes, l'intervention a permis de réduire la mortalité maternelle de 27 % par rapport aux changements observés dans le groupe contrôle.

**Tableau 3. Risques de mortalité maternelle selon les groupes et la période - essai QUARITE**

	Année 1 Décès p. 1000 (effectifs)	Année 4 Décès p. 1000 (effectifs)	Différence de risques p. 1000 (IC à 95 %)	Rapport de cotes (IC à 95 %)
Groupe Intervention	10,4 (445/43269)	6,8 (356/52662)	- 3,5 (- 2,3; - 4,7)	0,64 (0,57; 1,15)
Groupe Contrôle	8,1 (337/41655)	7,1 (381/53581)	- 1,0 (- 0,1; - 2,0)	0,88 (0,76; 1,16)

Les résultats bruts estimés avec l'approche DD, que ce soit en termes de différence de risques ou de rapport de cotes, montrent que l'intervention est efficace pour réduire la mortalité maternelle hospitalière dans un système de santé tel que celui du Mali ou du Sénégal. Même si l'impact mesuré semble faible, ces résultats seraient suffisamment importants pour convaincre un décideur ou une décideuse de mettre en œuvre un tel programme à l'échelle de son pays sur l'ensemble des hôpitaux. Toutefois, cette personne pourrait remettre en cause les résultats de l'étude qui ne prennent pas en compte le déséquilibre initial entre les groupes ni le caractère corrélé des données. Nous verrons dans les deux paragraphes suivants comment prendre en compte l'effet du groupe et l'effet grappe.

#### Prise en compte de l'effet du groupe

La figure 2 ci-dessus montre que le niveau de la mortalité maternelle à l'année 1 était plus élevé dans les hôpitaux du groupe d'intervention que dans le groupe contrôle. La diminution plus rapide de la mortalité dans le bras expérimental a ramené ensuite le taux de décès maternel à un niveau quasi identique à celui du groupe contrôle à l'année 4. Si nous avons comparé les

résultats de l'essai entre les deux groupes à l'année 4, nous aurions conclu à tort que l'intervention n'avait aucun effet sur la mortalité.

L'approche DD, qui analyse simultanément les changements observés dans les deux groupes, a permis de pallier ce problème, mais elle n'a pas contrôlé le déséquilibre initial entre les groupes. Comme la randomisation des hôpitaux était antérieure à l'inclusion des patientes, leurs caractéristiques n'étaient pas réparties de façon homogène entre les bras de l'essai.

**Tableau 4. Caractéristiques des patientes selon la période et le groupe - essai QUARITE**

	Année 1		Année 4	
	Intervention (n=4326 9)	Contrôle (n=41 655)	Intervention (n=52662)	Contrôle (n=53581)
Réside loin de l'hôpital (en dehors de la région)	2 242 (5,2)	509 (1,2)	1 115 (2,1)	854 (1,6)
Évacuée d'un autre établissement	11 644 (26,9)	9 384 (22,5)	15 382 (29,2)	13 097 (24,4)
Âge ≥ 35 ans	4 356 (10,1)	4 195 (10,1)	5 325 (10,1)	5 210 (9,7)
Nullipare	28 435 (65,7)	26 939 (64,7)	35 908 (68,2)	36 397 (67,9)
Antécédents de césarienne	3 112 (7,2)	2 782 (6,7)	5 196 (9,9)	4 920 (9,2)
Pathologie avant la grossesse	406 (0,9)	324 (0,8)	516 (1,0)	648 (1,2)
Aucune visite prénatale	4 221 (9,8)	4 535 (10,9)	4 780 (9,1)	5 250 (9,8)
Pathologie pendant la grossesse	3 976 (9,2)	3 401 (8,2)	3 897 (7,4)	4 834 (9,0)
Grossesse multiple	1768 (4,1)	1555 (3,7)	2 139 (4,1)	2 135 (4,0)

Dans notre cas, la proportion des femmes qui résidaient loin de l'hôpital

où elles avaient accouché était plus importante dans le bras expérimental que dans le groupe contrôle, et de ce fait, les évacuations sanitaires étaient plus fréquentes parmi les hôpitaux bénéficiant de l'intervention (Tableau 4). Étant donné la faible performance des filières de soins au Mali comme au Sénégal et les retards accumulés dans la prise en charge lors des évacuations, l'état de santé des femmes du bras expérimental était globalement plus précaire et leur risque de mortalité plus élevé. Le déséquilibre entre les groupes semble s'atténuer à l'année 4. On peut alors se poser la question du biais que cela peut entraîner dans l'analyse d'efficacité de l'intervention selon l'approche DD : est-ce que la diminution plus importante de la mortalité maternelle dans le bras expérimental ne s'explique pas tout simplement par la modification du profil des patientes dans ce groupe ?

Plusieurs méthodes statistiques permettent de répondre à cette question en prenant en compte les facteurs de confusion dans l'étude d'impact. Les modèles de régression logistique sont particulièrement adaptés à ce type d'analyse, car ils permettent d'ajuster la mesure de l'effet (DD ou ratio des rapports de cotes) à des caractéristiques individuelles ou collectives (de grappe), telles que le lieu de résidence ou le mode d'admission des patientes dans l'hôpital. Nous verrons dans le paragraphe suivant que les modèles de régression logistique généralisés permettent de prendre en compte, dans le même temps, l'effet grappe.

### Prise en compte de l'effet grappe

Puisque les patientes qui ont accouché dans le même hôpital ne sont pas totalement indépendantes les unes des autres, l'information rapportée par un hôpital de 1 000 femmes, par exemple, est donc moindre que s'il s'agissait de 1 000 patientes indépendantes ayant accouché dans différents centres. Il en résulte une perte de puissance dans l'analyse statistique qui doit être compensée par une augmentation du nombre de sujets à inclure. Par ailleurs, l'analyse des données recueillies doit recourir à des méthodes statistiques appropriées, sous peine d'obtenir des résultats abusivement significatifs. En effet, si l'on ne tenait pas compte de cet effet grappe, les étendues des intervalles de confiance s'en trouveraient faussement diminuées, car les tests statistiques ne prendraient pas en compte les variations inter- et intra-hospitalières de la mortalité maternelle. Différentes méthodes permettent

d'analyser correctement ce type de données parmi lesquelles l'utilisation de modèles mixtes ou marginaux. La commande *diff* du logiciel STATA (version 13.0), par exemple, permet d'estimer la DD en tenant compte de la structure hiérarchique des données, organisée en plusieurs niveaux imbriqués (niveau 1: patientes ; niveau 2: hôpital) et des éventuelles variables de confusions. La commande *xtlogit* du même logiciel permet d'estimer le ratio des rapports de cotes selon la même approche multivariée et multiniveaux.

**Tableau 5. Estimations brutes et ajustées des effets de l'intervention – essai QUARITE**

Méthode de mesure	Estimation brute (IC à 95 %)	Estimation ajustée* (IC à 95 %)
Différence de risques	-2,5 (- 3,5; - 1,5)	-2,5 (- 4,2; - 0,9)
Rapports de cotes	0,73 (0,61; 0,91)	0,85 (0,73; 0,98)

\* effet ajusté sur l'effet grappe et sur les caractéristiques des patientes et des hôpitaux

Après ajustement sur les variables de confusion et sur l'effet grappe, le ratio des RC est passé de 0,73 (estimation brute) à 0,85 (estimation ajustée), comme le montre le tableau 5. En d'autres termes, le déséquilibre de certaines caractéristiques individuelles, telles que le niveau d'accessibilité à l'hôpital, explique en partie la diminution plus importante de la mortalité maternelle dans le bras expérimental comparé au bras témoin, mais pas totalement. On peut donc attribuer une diminution relative de 15 % de la mortalité maternelle à l'intervention. De plus, après ajustement sur l'effet grappe, les étendues des intervalles de confiance sont plus importantes que dans les estimations brutes, quelle que soit la méthode de mesure utilisée (différence de risques ou rapport de cotes). On a donc perdu un peu de précision dans les mesures d'impact, mais les différences restent significatives compte tenu de la taille importante des échantillons.

## Analyse réflexive et conclusion

Nous avons vu à travers ce chapitre l'intérêt de mener un ECR en grappe pour évaluer l'impact des interventions en santé maternelle et comment

contrôler les trois principaux biais dans la mesure d'efficacité : l'effet du temps, l'effet du groupe et l'effet grappe. Il existe d'autres particularités des essais en grappe que nous aborderons dans la discussion sans toutefois les détailler. Ces particularités tiennent aux questions suivantes : l'expérimentation se fait-elle en aveugle ? Le consentement des sujets est-il nécessaire ? Les résultats sont-ils généralisables à d'autres contextes ?

Lorsque l'intervention que nous voulons tester prend la forme d'un programme d'activités ciblant des grappes d'individus (comme le programme d'amélioration de la qualité de soins dans l'essai QUARITE qui ciblait les professionnel-le-s de santé), l'inclusion des sujets se fait alors inévitablement en ayant connaissance du bras dans lequel il sera inclus. L'essai est dit « ouvert », contrairement à la plupart des essais individuels réalisés en « aveugle », c'est-à-dire sans avoir connaissance du bras dans lequel les sujets sont répartis aléatoirement. Cette situation est potentiellement source de biais qui peuvent être induits par des comportements éventuellement différents entre les groupes.

La question du recueil de consentement pour participer à l'étude est également importante. Si le consentement du ou de la responsable de grappe est incontournable, celui des sujets n'est pas nécessairement requis. Il est possible d'informer les individus susceptibles de participer à l'étude, sans toutefois obtenir un consentement signé de leur part. Dans l'essai QUARITE, par exemple, nous avons obtenu le consentement du directeur d'hôpital et du chef de service de la maternité. En revanche, nous ne l'avons pas fait pour les femmes ayant accouché dans chaque hôpital participant à l'étude, puisque nous n'intervenions pas directement sur elles et que les données collectées anonymisées étaient issues des registres hospitaliers existants, sans interview auprès des femmes.

Une autre singularité de l'essai en grappe tient au fait que le contexte dans lequel l'intervention est mise en œuvre peut varier considérablement d'une grappe à l'autre. Si l'on connaît à l'avance les caractéristiques des grappes qui ont un effet sur le critère de jugement principal, il est d'ailleurs recommandé de stratifier la randomisation sur ces critères. Dans le cas de l'essai QUARITE, nous avons randomisé des hôpitaux en fonction du niveau de soins, car les données disponibles préliminaires nous indiquaient que les taux de mortalité étaient différents entre les hôpitaux de district, les hôpitaux régionaux et les hôpitaux de la capitale. L'analyse des résultats selon le niveau de soins montrait que l'intervention était particulièrement

efficace dans les hôpitaux de district et peu efficace dans les hôpitaux régionaux. Nous avons donc conclu que les bénéfices de l'intervention devaient s'appliquer uniquement aux hôpitaux de première ligne et que d'autres recherches seraient nécessaires pour confirmer ou non l'utilité d'un tel programme dans des hôpitaux de niveau supérieur.

Même s'ils sont complexes à mettre en œuvre, les essais contrôlés randomisés sont les études d'impact en santé maternelle et périnatale qui fournissent le niveau de preuve le plus élevé s'ils sont menés dans les règles de l'art. Pour des raisons logistiques, organisationnelles et parfois éthiques, il est souvent plus pertinent de randomiser des unités « sociales » ou grappes que des individus. Ces essais en grappes entraînent des défis méthodologiques qui ne sont pas insurmontables à condition de les avoir pris en compte dans la mise en place de l'étude et dans les analyses. On doit également garder à l'esprit que, dans ce type d'étude, l'unité de randomisation est différente de l'unité d'analyse, c'est-à-dire l'unité qui sera prise en compte pour l'interprétation des résultats ou l'analyse d'impact. Mais le niveau de preuve d'efficacité d'une intervention n'est pas l'unique critère de choix pour les décideuses et décideurs. Le contexte et les conditions de mise en œuvre sont aussi des éléments essentiels pour concevoir des politiques publiques plus efficaces. Les futures recherches devraient permettre également de comprendre comment et pourquoi cette intervention produit des effets positifs sur la santé des mères et des nouveau-nés. Les essais hybrides combinant à la fois les essais contrôlés en grappe et l'analyse d'implantation sont les plus adaptés pour répondre simultanément aux questions d'efficacité et de mise en œuvre.

## Références clés

Campbell, M. K., Mollison, J., Steen, N., Grimshaw, J. M. et Eccles, M. (2000) Analysis of cluster randomized trials in primary care: a practical approach. *Family Practice*, 17(2), 192-196.

*Un article qui présente très clairement la méthode des essais randomisés en grappe.*

Dumont, A., Fournier, P., Abrahamowicz, M., Traoré, M., Haddad, S. et Fraser, W. (2013). Quality of care, risk management, and technology in obstetrics : a cluster-randomized trial of a multifaceted intervention to reduce

hospital-based maternal mortality in Senegal and Mali (QUARITE). *Lancet*, 382(9887), 146 – 157.

*Les résultats de cette étude réalisée au Sénégal et au Mali ont été utilisés pour illustrer ce chapitre.*

Chaillet, N., Dumont, A., Abrahamowicz, M., Pasquier, J. C., Audibert, F., Monnier, P., ... et Fraser, W. D. (2015). A cluster-randomized trial to reduce cesarean delivery rates in Quebec. *New England Journal of Medicine*, 372(18), 1710-1721.

*Les résultats de cette étude réalisée au Canada permettent d'illustrer les défis méthodologiques liés aux essais en grappe.*

## Références

Dumont, A., Fournier, P., Abrahamowicz, M., Traoré, M., Haddad, S. et Fraser W. (2013). Quality of care, risk management, and technology in obstetrics : a cluster-randomized trial of a multifaceted intervention to reduce hospital-based maternal mortality in Senegal and Mali (QUARITE). *Lancet*, 382(9887), 146–157

Stanton, C. K., Newton, S., Mullany, L. C., Cofie, P., Agyemang C. T., Adiibokah, E., ... et Gyapong, J. (2013). Effect on postpartum hemorrhage of prophylactic oxytocin (10 IU) by injection by community health officers in Ghana: A community-based, cluster-randomized trial. *PLoS Medicine*, 10(10), e1001524.

doi: 10.1371/journal.pmed.1001524

Villar, J., Ba'aqeel, H., Piaggio, G., Lumbiganon, P., Belizán, J. M, Farnot, U., ... et Garcia, J. (2001). WHO antenatal care trial research group. *Lancet*, 357(9268), 1551-64.

## Résumé / Abstract / Resumen

Les essais contrôlés randomisés en grappe permettent d'évaluer l'impact des interventions en santé avec un niveau de preuve très élevé. Si les essais randomisés en grappes offrent plusieurs avantages dans les pays à faibles ressources par rapport aux essais individuels, ils présentent quelques défis



méthodologiques qui tiennent au fait que l'unité d'analyse est différente de l'unité de randomisation ou d'intervention. Ces défis tant sur le plan de la mise en œuvre de l'intervention que de l'analyse statistique sont présentés à partir d'une expérience réalisée au Mali et au Sénégal (essai QUARITE) pour réduire la mortalité maternelle et néonatale hospitalière. Nous verrons comment l'approche « différence des différences » permet de mesurer l'effet l'intervention tout en tenant compte de l'effet du temps qui joue souvent favorablement sur les indicateurs de santé en dehors de toute intervention extérieure. Les modèles statistiques appropriés sont présentés pour pouvoir ajuster l'effet de l'intervention sur les différences qui existent entre les groupes expérimentaux avant le démarrage du programme à tester et sur l'effet de groupe propre aux essais en grappe.

\*\*\*

Cluster-randomised trials are used to assess the impact of health interventions with a very high level of evidence expected. While cluster randomized trials offer several advantages in low-resource countries compared to individual trials, they present some methodological issues because the unit of analysis is different from the unit of randomization and intervention. These challenges, both in terms of the intervention implementation and the statistical analysis, are presented based on a trial conducted in Mali and Senegal (QUARITE trial) to reduce maternal and neonatal mortality in hospitals. We will see how the difference-in-differences approach measures the effect of the intervention taking into account the secular trends, which often plays favorably on health indicators outside of any external intervention. The appropriate statistical models are presented in order to adjust the effect of the intervention on baseline case-mix and on the cluster effect which is specific to this type of trial.

\*\*\*

Los ensayos controlados aleatorios grupales permiten evaluar el impacto de las intervenciones de salud con un nivel muy alto de pruebas. Aunque los ensayos aleatorios grupales ofrecen varias ventajas en los países de bajos recursos con respecto a los ensayos individuales, presentan algunos desafíos metodológicos debido al hecho de que la unidad de análisis es diferente de la unidad de aleatorización o de intervención. Estos desafíos, tanto en la implementación de la intervención como en el análisis estadístico, se presentan sobre la base de un experimento realizado en Malí y Senegal

(ensayo QUARITE) para reducir la mortalidad materna y neonatal en los hospitales. Veremos cómo el enfoque de « diferencia de diferencias » permite medir el efecto de la intervención teniendo en cuenta el efecto del tiempo, que a menudo tiene un impacto positivo en los indicadores de salud fuera de cualquier intervención externa. Se presentan modelos estadísticos apropiados para ajustar el efecto de la intervención sobre las diferencias entre los grupos experimentales antes del inicio del programa de prueba y sobre el efecto del grupo específico de los ensayos de grupos.

\*\*\*

**Alexandre Dumont** est gynécologue-obstétricien et épidémiologiste. Il a mené des essais contrôlés en grappe en santé maternelle et périnatale au Canada, au Sénégal, au Mali et au Burkina Faso. Il a reçu le prix « Birthday Prize » de la Commission européenne pour l'essai QUARITE qui a permis de réduire la mortalité maternelle et néonatale dans les hôpitaux de référence du Mali et du Sénégal.

### Citation

Alexandre Dumont. (2019). L'évaluation de la pérennité. Un exemple en santé maternelle et infantile. In *Évaluation des interventions de santé mondiale. Méthodes avancées*. Sous la direction de Valéry Ridde et Christian Dagenais, pp. 263-280. Québec : Éditions science et bien commun et Marseille : IRD Éditions.

Dumont Alexandre. (2019).

Les essais randomisés en grappe : un exemple en santé maternelle et infantile.

In : Ridde Valéry (ed.), Dagenais C. (ed.) Évaluation des interventions de santé mondiale : méthodes avancées.

Québec (CAN) ; Marseille : Ed. Science et Bien Commun ; IRD, 265-282. ISBN 978-2-7099-2766-6