UCLouvain

FACULTE DES SCIENCES

Faculté des sciences

Essais croisés, essais stepped wedge et essais N of 1 : techniques d'analyse et calculs de taille d'échantillon

Auteur·es : STETENFELD Emilie

Promoteur·rices : LEGRAND Catherine

Lecteur·rices : BUGLI Céline

Année académique : 2023-2024

Intitulé du master et de la finalité : Master [120] en statistique, orientation biostatistiques, à finalité spécialisée

Remerciements

Je tiens tout d'abord à remercier ma promotrice, madame LEGRAND Catherine. Elle m'a soutenue tout au long de la création de ce travail ainsi que durant mes deux années de master. Son aide a été précieuse. J'ai su construire et améliorer ce mémoire grâce à ses conseils et ses nombreuses relectures effectuées tout au long de la création de ce projet. Ses encouragements et son soutien m'ont aidée à garder le moral durant cette dernière année de master.

Merci également à ma famille et à mes proches de m'avoir soutenue pendant ces deux années qui ont été particulièrement éprouvantes. Ils ont toujours été là pour me remonter le moral et me soutenir dans les moments difficiles.

Je tiens également à remercier l'ensemble des professeurs du master qui sont particulièrement impliqués dans l'apprentissage et l'accompagnement des étudiants. Merci également aux assistants qui se sont toujours montrés disponibles en cas de difficultés.

Table des matières

1 Introduction

L'idée des essais cliniques est apparue à partir du XVIII^{ème} siècle. Un essai clinique est défini comme une étude prospective, réalisée sur l'Homme, comparant les effets et la valeur d'un (ou de plusieurs) traitement(s) par rapport à un témoin (Friedman, Furberg, DeMets, Reboussin, & Granger, 2015). De nombreuses années plus tard, Fisher a été l'un des premiers à introduire le concept de la randomisation en 1926 (*ibidem*). A partir des années 1930, les essais cliniques randomisés ont fait leur apparition dans l'univers scientifique (*ibidem*). Au fur et à mesure des années, ces essais sont devenus le type d'essais le plus utilisé pour évaluer et étudier différents types de traitements dans le secteur de la médecine (*ibidem*). Avec le temps, de nombreuses techniques, méthodes et analyses se sont développées permettant l'utilisation d'essais cliniques pour analyser des données plus complexes.

Les essais cliniques sont composés de différentes phases (Friedman, Furberg, DeMets, Reboussin, & Granger, 2015). La phase I permet d'obtenir des informations en ce qui concerne la tolérabilité, la pharmacocinétique[1] et la pharmacodynamique[2] du (des) traitement(s). Cette dernière est réalisée sur un très petit nombre de volontaires sains. Elle apporte une information sur la toxicité immédiate et permet donc d'avoir une idée de la dose à utiliser pour les phases suivantes. De plus, elle apporte des informations en ce qui concerne l'action du médicament. La phase II est réalisée sur un petit nombre d'individus atteints de la maladie d'intérêt. Cette phase permet d'obtenir des informations en ce qui concerne les propriétés biologiques et l'activité d'un traitement. Elle est souvent utilisée pour obtenir des informations nécessaires pour la mise en place de la phase III, comme par exemple une estimation de la probabilité de succès. Les phases III et IV sont les phases qui sont considérées tout au long de ce travail. Ces phases sont réalisées sur un grand nombre d'individus atteints de la maladie d'intérêt. Elles permettent d'obtenir une estimation de l'effet traitement. La phase III permet d'étudier l'effet traitement. La phase IV peut être considérée comme le prolongement de la phase III. Elle permet d'étudier les potentiels effets secondaires du traitement sur du long terme, après la mise sur le marché du traitement.

Les essais cliniques randomisés et réalisés sur un groupe d'individus suffisamment large constituent la meilleure méthode pour déterminer quel traitement est le plus efficace et ainsi améliorer la Santé Publique (Friedman, Furberg, DeMets, Reboussin, & Granger, 2015). Les essais contrôlés randomisés (RCT de l'anglais 'Randomized Controlled Trials') sont les premières études de type essais cliniques à avoir vu le jour. Ces études sont caractérisées par la comparaison du groupe traitement au groupe contrôle, les individus étant répartis aléatoirement dans chacun des groupes (Twisk, 2021). Le groupe traitement correspond à des individus qui reçoivent le nouveau traitement étudié. Le groupe contrôle est constitué d'individus qui reçoivent un traitement de contrôle. Généralement, ce traitement peut être un placebo, un traitement actif ou l'absence totale de traitement (aucune administration). Les

[1] L'étude du devenir du médicament dans l'organisme (Les éditions Le Robert et Google, 2014)
[2] L'étude de l'action du médicament sur l'organisme (Les éditions Le Robert et Google, 2014)

RCT en groupes parallèles font partie des types d'études RCT souvent utilisés en médecine. Dans ces études, les individus sont répartis aléatoirement dans chacun des groupes, de contrôle et de traitement. Ces derniers reçoivent alors soit le traitement de contrôle soit le nouveau traitement et ils sont ensuite suivis pendant la même période en 'parallèle'. L'analyse des données consiste alors à comparer les valeurs obtenues pour une variable d'intérêt (réponse ou « outcome ») pour chacun des groupes, de traitement et de contrôle. L'objectif est d'évaluer s'il existe au niveau de la population une vraie différence entre les valeurs obtenues pour le traitement et celles obtenues pour le contrôle, appelée l'effet traitement. Traditionnellement, cette différence est évaluée sous forme de tests d'hypothèses (Friedman, Furberg, DeMets, Reboussin, & Granger, 2015). Le principe est de tester l'hypothèse nulle, qui considère un effet traitement égal à 0 donc aucune différence entre le traitement et le contrôle, et de ne pas la rejeter ou de la rejeter. Dans ce dernier cas, l'hypothèse nulle est rejetée en faveur de l'hypothèse alternative qui considère un effet traitement différent de 0 donc une réelle différence entre le traitement et le contrôle.

Depuis de nombreuses années, il est reconnu que la randomisation est un processus d'une grande importance dans les essais cliniques. Ce concept a été introduit par Fisher en 1926 (Friedman, Furberg, DeMets, Reboussin, & Granger, 2015). Elle est définie comme le processus au cours duquel les participants sont assignés aléatoirement à un groupe de traitement et/ou un groupe contrôle (Kang, Ragan, & Park, 2008). La randomisation est un processus important pour plusieurs raisons. Premièrement, elle permet d'homogénéiser au mieux les différents groupes et éviter ainsi les différences systématiques. Par exemple, si une étude porte sur un temps de marche lors d'une randonnée et qu'un des groupes est constitué principalement de personnes en moyenne plus âgées par rapport aux autres groupes, les résultats pourraient être influencés par cette répartition (Kang, Ragan, & Park, 2008). Un second exemple pourrait être une étude de la pression artérielle des individus. Si les participants recrutés dans un groupe sont tous des personnes âgées ou des enfants comparés aux autres groupes, alors les groupes ne sont pas comparables. En effet, les personnes âgées ont une tension artérielle généralement plus élevée et les enfants une tension plus basse. Deuxièmement, la randomisation permet de ne pas connaître la répartition des groupes, de mettre 'en aveugle' les participants, les chercheurs ou tout autre intervenant à l'étude. C'est-à-dire que les participants de l'étude ne savent pas quel traitement ils reçoivent et/ou les chercheurs n'ont pas connaissance de quel traitement est administré à chacun des patients. A savoir que le fait de connaître la répartition des individus aux différents groupes pourrait influencer les résultats (Kang, Ragan, & Park, 2008). Cependant, cette mise en aveugle n'est pas toujours possible même lorsque les individus ont été répartis aléatoirement. Ce dernier point sera expliqué plus en détails par la suite, notamment dans le cas des essais *stepped wedge*.

Un dernier point très important à considérer lors de la mise en place des essais cliniques est le calcul de la taille d'échantillon. Dans les essais cliniques en général, la taille de l'échantillon relève du domaine de l'éthique (Senn S. , 2002). Cette taille d'échantillon est en réalité un

équilibre entre 'trop' et 'trop peu' d'individus recrutés. En effet, sélectionner trop peu d'individus présents dans l'étude ne permettrait pas de détecter de façon statistiquement significative une différence au niveau de la population entre les traitements étudiés. Tandis qu'exposer trop d'individus aux risques éventuels des traitements serait éthiquement inacceptable. La taille d'échantillon doit donc permettre de mettre en évidence une différence entre les traitements étudiés avec un nombre d'individus minimum. Le calcul de la taille d'échantillon permet en réalité de contrôler l'erreur de type I et de type II pour le test d'hypothèse comparant les deux groupes.

Au cours des années, différents types d'essais cliniques ont vu le jour. Récemment, l'intérêt porté à deux types particuliers d'essais cliniques, nommés essais *stepped wedge* et essais *N of 1*, a augmenté. Les essais *stepped wedge* et les essais *N of 1* sont deux types d'essais dérivés de ceux connus depuis de nombreuses années, appelés les essais croisés. Ces trois types d'études ont la particularité d'administrer l'ensemble des traitements étudiés à chacun des individus de l'étude. Les essais stepped wedge et les essais *N of 1* sont des études qui ont été développées pour être mises en place dans des situations particulières, comme par exemple dans le cas de l'étude des maladies rares pour les essais *N of 1*.

Si les méthodes d'analyse des essais croisés sont connues et développées depuis plusieurs années, elles ont néanmoins besoin d'être adaptées afin de pouvoir être utilisées pour des types d'études plus récentes comme les essais *stepped wedge* et les essais *N of 1*. Étant donné que l'intérêt porté pour ces deux designs est en augmentation depuis peu, il est nécessaire de résumer et regrouper les différentes techniques et analyses présentes dans la littérature scientifique. Les essais *stepped wedge* et *N of 1* vont être présentés en parallèle des essais croisés. Ces derniers sont des essais plus connus pour lesquels de nombreuses techniques d'analyse sont déjà utilisées par un grand nombre de scientifiques.

Dans un premier temps, une présentation globale de chacun des types d'études va d'abord être réalisée. Cette première partie permettra de poser les bases théoriques afin de poursuivre et comprendre la suite des analyses. Dans un deuxième temps, des techniques d'analyse seront présentées. Cette section présente différentes techniques d'analyse des données récoltées suite à la mise en place des essais croisés ainsi que des essais *stepped wedge* et *N of 1*. Tout au long de ce travail, la variable d'intérêt qui permet de mesurer l'effet traitement sera considérée comme continue et normalement distribuée. Il est nécessaire de garder à l'esprit que les techniques présentées dans ce travail peuvent être étendues aux cas où cette variable n'a pas une distribution normale. Enfin, différentes méthodes pour calculer la taille d'échantillon nécessaire à la mise en place de ces types d'études seront présentées. Ces dernières sont spécifiques à chacun des designs. Certaines de ces techniques ont été développées très récemment.

2 Présentation des différents types d'études

2.1 Essais croisés

Ce travail, et particulièrement les sections concernant les essais croisés, a été rédigé à l'aide de l'ouvrage « Cross-over Trials in Clinical Research », écrit par le statisticien Stephen SENN (Senn S. , 2002).

Les essais croisés (ou *cross-over* en anglais) sont des essais caractérisés par l'administration de plusieurs traitements sous forme de séquences à un même individu. Idéalement, les individus de l'étude reçoivent l'ensemble des traitements dans un ordre différent. Le but principal de ce type d'essais est de mesurer et de comparer les effets de chacun des traitements étudiés.

Ces études sont caractérisées par des séquences qui sont-elles même constituées de plusieurs périodes. Une séquence correspond à l'ordre dans lequel les patients vont recevoir les traitements. En effet, si deux traitements sont étudiés, les patients peuvent recevoir les traitements dans un ordre, traitement A puis traitement B ou dans l'autre, traitement B puis traitement A. Ces séquences sont attribuées aléatoirement à chacun des patients. Ce dernier point est détaillé à la fin de cette section.

Les essais croisés sont des essais réalisés pour étudier l'efficacité de différents traitements au cours de différentes périodes. Ces traitements ont pour objectif de diminuer les effets d'une pathologie et non d'éliminer cette dernière. En effet, l'objectif est de comparer l'efficacité de deux traitements administrés successivement à un patient dont le statut ne change pas au cours du temps. Si un des traitements guérit le patient alors l'état du patient change, en passant d'un statut de malade à un statut de non malade (guérison). Dans ce cas, la comparaison des traitements n'est plus envisageable. De plus, l'administration d'un second traitement à un patient guéri ne présente plus d'intérêt. Ce design est donc particulièrement utilisé dans le cas des maladies chroniques. Les maladies chroniques, appelées aussi maladies non transmissibles, sont définies par l'Organisation Mondiale de la Santé comme des maladies de longues durée, à évolution lente (OMS, 2023). Ces maladies étant rarement curables, l'objectif est de modérer les effets néfastes de celles-ci à l'aide d'un traitement en particulier. L'asthme, l'hypertension ou encore le rhumatisme sont des exemples de situations dans lesquelles les essais croisés sont fréquemment utilisés.

La *figure 1*, représentée ci-dessous, illustre la structure générale des essais croisés :

	Traitements	
Séquence n°1	A	B
Séquence n°2	B	A

Figure 1 : Illustration de la structure des essais croisés

Une étude bien connue de la littérature qui a eu une grande importance dans l'histoire des essais croisées est l'étude des isomères optiques sur la durée du sommeil des détenus à

Kalamazoo, rapportée par Cushny et Peebles (Cushny & Peebles , 1905). Au cours de cette étude, 3 traitements ainsi qu'une période de contrôle (sans traitement) ont été administrés à chacun des individus. L'objectif final était d'étudier l'effet individuel de chaque traitement en administrant chacun d'eux à l'ensemble des individus. Ces données ont notamment participé à l'inauguration de l'ère statistique moderne et ont été utilisées dans de nombreux manuels influents sur le sujet.

Pour réaliser l'étude présentée ci-dessus, un essai en groupes parallèles aurait pu être envisagé. Dans ce cas, il aurait fallu recruter un certain nombre de patients, les répartir aléatoirement en 4 groupes et attribuer un traitement par groupe. Chacun des patients n'aurait alors reçu qu'un seul traitement.

Toutefois, l'utilisation des essais croisés comparés aux essais en groupes parallèles offre plusieurs avantages.

Un premier avantage est l'utilisation d'un nombre moins élevé d'individus. Étant donné que plusieurs traitements sont attribués à un même patient, le nombre d'individus à recruter comparé à un essai en groupes parallèles pour un même nombre d'observations est moindre. Puisque dans les essais croisés le patient reçoit tous les traitements étudiés, les mesures réalisées chez un même patient sont répétées. Chaque patient est alors utilisé comme son propre témoin. La comparaison des différents traitements chez un même patient dépend alors de la variation intra-patient, et non plus de la variation inter-patient comme c'est le cas pour les essais en groupes parallèles. A savoir que la variance intra-patient est généralement moindre que la variance inter-patient (Elbourne, et al., 2002). Un deuxième avantage est donc que les essais croisés permettent d'obtenir des résultats plus précis que les RCT et ce même si les essais croisés sont constitués d'un nombre moins élevé d'individus (Elbourne, et al., 2002).

Cependant, les essais croisés comportent également quelques inconvénients. Certains d'entre eux sont liés directement aux patients comme par exemple, le problème du « drop-out ». Le drop-out survient lorsque le patient interrompt son traitement avant la fin de l'essai. Ce risque existe également dans les essais en groupes parallèles. Cependant, l'effet du drop-out dans les essais croisés peut être plus conséquent, comme par exemple, lorsque le patient interrompt la prise dès le premier traitement. Un autre inconvénient lié au patient est le désagrément que peut entraîner la prise de plusieurs traitements différents à la suite. A noter que ce point peut devenir un avantage si le patient souhaite tester plusieurs traitements afin d'acquérir une expérience personnelle des effets de chacun.

Les indications et les conditions de mise en place d'un essai croisé peuvent également constituer un point négatif de ce type d'études. En effet, pour réaliser une étude de type essai croisé, il est nécessaire d'étudier une maladie stable dans le temps. Les patients faisant l'objet de l'étude ne doivent pas avoir un risque élevé de mourir pendant la durée d'observation ou encore, un risque de voir leur état de santé se détériorer ou même s'améliorer. L'état du patient doit rester stable afin de pouvoir comparer les effets réels des traitements reçus au

cours des différentes périodes. De nombreuses maladies ne se prêtent donc pas aux conditions pour réaliser une étude à essais croisés.

Un dernier inconvénient, et non des moindres, est le *carry over* (appelé en français effets résiduels, effets de rémanence ou encore effets de report). Ce phénomène est défini comme la persistance, physique ou en termes d'effets, d'un traitement appliqué lors d'une période antérieure. Cela signifie qu'un traitement qui est donné lors d'une première période pourrait modifier les effets d'un deuxième traitement donné à la période suivante. L'effet mesuré à la deuxième période n'est alors plus le résultat du deuxième traitement uniquement mais bien d'une interaction entre les deux traitements. Au-delà du fait qu'il est difficile d'interpréter et d'analyser les résultats lorsque le *carry over* est présent, il est également très difficile de le détecter. Une période de *wash-out* peut être ajoutée afin de limiter au maximum cet effet *carry over*. La période de *wash-out* est une période au terme de laquelle l'effet du traitement administré précédemment est considéré comme ayant disparu. Elle est considérée comme passive si le patient ne reçoit aucun traitement au cours de cette période et comme active si un traitement est administré.

Dans les essais croisés, la randomisation ne se fait pas au niveau des groupes comme c'est le cas dans les RCT mais plutôt au niveau des séquences. Par exemple, dans un essai croisé de deux traitements A et B, les patients sont randomisés soit dans une première séquence où ils recevront le traitement A puis B, soit dans une séquence où ils recevront B puis A.

Le calcul de la taille de l'échantillon étudié est également un processus primordial. Le choix de la taille d'échantillon dans les essais croisés s'avère être plus complexe que dans les essais en groupes parallèles. Le calcul de la taille d'échantillon des essais croisés va dépendre, en partie, de la variance de la variable réponse. Ce point sera développé par la suite (voir point 4.1).

2.2 Essais *stepped wedge*

Les essais *stepped wedge* (appelés aussi en français essais à permutations séquentielles) sont définis comme des essais croisés unidirectionnels dans lesquels plusieurs groupes passent d'une situation de contrôle à une situation de traitement à des moments différents (Twisk, 2021). Le terme unidirectionnel signifie qu'un groupe peut passer d'une situation A à une situation B, mais il ne repassera jamais de B à A.

Plusieurs précisions sont à faire sur les essais *stepped wedge*. Ces derniers, comme défini ci-dessus, concernent des groupes d'individus. Ces groupes correspondent généralement à un ensemble d'individus repris dans des institutions (hôpitaux, centre de soins, écoles…). En réalité, les essais *stepped wedge* sont des essais dans lesquels un traitement est administré en même temps à l'entièreté d'une institution, d'un centre ou d'une école par exemple, donc par définition à un ensemble d'individus. Une deuxième notion importante est la notion de traitement utilisée dans le contexte des essais *stepped wedge*. Dans ce type d'essais, le terme *intervention* est généralement préféré par les chercheurs plutôt que le terme *traitement*. Il est vrai que les essais *stepped wedge* sont le plus souvent utilisés pour mettre en place une intervention plutôt qu'un traitement. L'implémentation d'une nouvelle intervention

chirurgicale, d'une intervention de prévention des chutes ou encore d'un processus d'accompagnement des douleurs chroniques sont des exemples de situations dans lesquelles un essai *stepped wedge* peut être mis en place. Dans ces différents exemples, ce n'est effectivement pas un traitement qui est étudié mais plutôt une intervention. Cependant, pour faciliter la compréhension et harmoniser la rédaction de ce travail, le terme *traitement* sera utilisé dans le cas des essais *stepped wedge* comme un terme général englobant également le terme *intervention*.

La *figure 2* donne un exemple de la structure des essais *stepped wedge* dans lequel 3 groupes passent progressivement de l'absence de traitement (O) au traitement (I).

	Temps 0	**Temps 1**	**Temps 2**	**Temps 3**
Groupe n°1	O	I	I	I
Groupe n°2	O	O	I	I
Groupe n°3	O	O	O	I

Figure 2 : Illustration de la structure des essais Stepped Wedge

Tous les groupes sont recrutés au même moment et ils sont suivis durant les mêmes périodes comme le montre la *figure 2.* La première période est commune pour tous les groupes et correspond à une période de contrôle, donc sans administration du traitement étudié. La différence entre les groupes d'individus est que la période à laquelle le traitement est mis en place est différente en fonction des groupes. Comme montré dans le tableau ci-dessus, la mise en place du traitement dans le groupe n°1, qui peut être pour rappel un hôpital ou une école par exemple, est réalisée au temps n°1. Tandis que la mise en place du traitement est réalisée au temps n°2 pour le groupe n°2. L'analyse des données est ensuite réalisée en comparant les individus de la période de contrôle à ceux de la période de traitement.

Une des premières utilisations du design *stepped wedge* est apparue dans les années 1980 en Gambie. Le vaccin ayant démontré son efficacité contre l'hépatite B, l'objectif était d'étudier l'efficacité du vaccin sur les maladies chroniques du foie. Afin d'étudier les bénéfices à long terme, le vaccin a été déployé aux nourrissons de manière séquentielle dans les différentes zones du pays. Cette administration progressive a conduit au final à une protection complète du pays. L'utilisation de ce type d'études est en augmentation. Des études au design *stepped wedge* ont été mises en place dans le cas de différentes maladies comme le VIH ou encore des cancers (Hemming K. H., 2015).

Une revue systématique, réalisée en 2006, a mis en évidence deux situations particulières dans lesquelles les essais *stepped wedge* sont principalement utilisés (Brown & Lilford, 2006). La première est lorsqu'il existe une conviction sur l'efficacité du traitement et non pas seulement une croyance. Par exemple, il est inacceptable d'envisager l'administration d'un vaccin, pour lequel il existe une conviction que son effet sera plus bénéfique que néfaste, uniquement chez certains individus, comme ce serait le cas dans un essai en groupes parallèles. Dans cette situation, les essais *stepped wedge* sont donc préférables puisqu'ils

présentent l'avantage d'administrer le traitement progressivement à l'ensemble des individus.

Ensuite, il pourrait exister des contraintes, comme des contraintes logistiques. Une contrainte logistique pourrait être, par exemple, l'impossibilité de mettre en place le traitement au même moment dans chacun des groupes (Federico, et al., 2022). Par exemple, si le traitement constitue la mise en place d'un vaccin, il faudra du temps pour obtenir un nombre de doses suffisant pour tous les groupes. Ce phénomène peut également apparaître lorsqu'un traitement qui demande une longue et complexe préparation du personnel soignant et des patients de l'étude est mis en place. Dans ce-cas, l'utilisation d'un essai *stepped wedge* permet de contourner ces contraintes logistiques éventuelles.

Le fait que les essais *stepped wedge* sont réalisés sur des groupes d'individus offre un autre avantage. En effet, lorsque les individus sont sélectionnés un à un, il est possible que certains individus soient sélectionnés plutôt que d'autres. Un exemple serait d'imaginer une situation dans laquelle un médecin généraliste est chargé de sélectionner des individus parmi ses patients pour réaliser un essai clinique. Le médecin connaissant bien ses patients sera peut-être plus tenté de sélectionner des jeunes individus en bonne santé ou encore des individus habitant les environs. En sachant que des individus plus âgés ou des personnes habitant plus loin seraient peut-être moins susceptibles de réagir au traitement ou encore de moins suivre rigoureusement l'étude. Ce biais s'appelle un biais de sélection. Les individus ne sont plus choisis au hasard mais bien parce qu'ils semblent particulièrement intéressants pour l'étude. Les études *stepped wedge* permettent donc d'éviter ce biais de sélection, puisque dans ces dernières, les groupes d'individus (hôpitaux ou centres de soins par exemple) reçoivent tous le traitement.

Plusieurs inconvénients sont à souligner. La durée des essais de type *stepped wedge* peut être plus longue, comparée à celle des essais croisés ou en groupes parallèles. La mise en place de ces essais peut également engendrer quelques complications pratiques. Ces dernières peuvent être, par exemple, le maintien d'une bonne coopération et d'un engagement des groupes qui doivent passer au traitement à un moment précis (calendrier de randomisation) (Hemming K. H., 2015) ou encore la mise en aveugle des évaluateurs. Cette dernière est particulièrement importante puisque dans les essais *stepped wedge*, selon le type de traitement, il peut s'avérer difficile de mettre en aveugle les participants ou les personnes impliquées dans l'étude. En effet, un exemple simple pourrait être la mise en place d'un soutien émotionnel dans le cas des douleurs chroniques (un sujet particulièrement étudié à l'aide des essais *stepped wedge*). Dans ce cas précis, le personnel soignant suivra une formation spécifique afin d'implémenter l'aide aux individus via des thérapies, par exemple. Ils seront tous conscients du passage de l'état de contrôle à l'état de traitement. Dans ce cas, la mise en aveugle du personnel soignant et des patients devient impossible. Les personnes qui récoltent les résultats doivent alors être mises en aveugle. Cela permet d'éviter le plus possible le biais d'information pouvant survenir si l'évaluateur a connaissance du statut du

patient (Brown & Lilford, 2006). Mais à nouveau, au vu du design des essais *stepped wedge*, mettre en aveugle l'évaluateur peut également être compliqué.

Comme pour les essais croisés, la randomisation dans le cas des essais *stepped wedge* est un processus important. Dans les essais *stepped wedge*, ce ne sont pas les individus qui sont randomisés mais les groupes d'individus (exemple les hôpitaux ou les écoles). La randomisation détermine le moment où le groupe passe de l'état de contrôle à l'état de traitement.

L'intérêt particulier pour les essais *stepped wedge* est récent. De nombreux auteurs se sont intéressés et ont développé des méthodes afin de trouver une taille d'échantillon optimale. Il existe donc de nombreuses approches pour calculer la taille d'échantillon des essais *stepped wedge* (Martin, Taljaard, Girling, & Hemming, 2016). Ce calcul dépend de plusieurs facteurs comme par exemple, le nombre de périodes de l'étude, le nombre de groupes repris dans l'étude et la corrélation intra-groupe. Ce point sera développé plus tard dans ce travail (voir point 4.2).

2.3 Essais *N of 1*

Les essais *N of 1* (appelés en français les études à cas unique) sont définis comme étant des essais dans lesquels un patient reçoit à plusieurs reprises deux traitements, ou plus, à des moments distincts (Senn S. , 2019). Le but premier de ces essais est d'obtenir des informations sur l'effet traitement spécifique à un patient. En effet, étant donné que le patient reçoit plusieurs fois plusieurs traitements, il est possible de mettre en évidence le traitement qui lui correspond le mieux. Ils peuvent être vus comme de multiples essais croisés réalisés sur un même patient (Richard & Naihua, 2019).

Cependant, depuis quelques années, certains scientifiques ont eu l'idée d'utiliser les essais *N of 1* sur plusieurs individus. Dans ce deuxième contexte, le but est d'étudier, sur base des observations récoltées sur chacun des patients, l'effet traitement (Senn S. , 2019). Il n'est alors plus question d'étudier comment chacun des traitements agit sur chaque individu mais d'étudier quel traitement est le plus bénéfique par rapport à l'ensemble des individus (Senn S. , 2024).

La *figure 3* permet de mieux visualiser la structure d'un essai *N of 1* (Richard & Naihua, 2019). Elle illustre l'administration de deux traitements, A et B, répartis en trois blocs constitués chacun de deux périodes.

Bloc	**1**	**1**	**2**	**2**	**3**	**3**
Période	1	2	3	4	5	6
Traitement	A	B	B	A	A	B

Figure 3 : Illustration de la structure des essais N of 1 sur un seul patient

Bien que les essais *N of 1* aient été proposés en 1953 par Hogben et Sim, ce n'est que 30 ans plus tard que ce type d'essais a réellement fait son entrée dans la littérature médicale (Richard

& Naihua, 2019). Une des premières études à utiliser ce design a été réalisée dans le but d'identifier le meilleur traitement contre l'asthme pour un patient donné (Guyatt, et al., 1986). Le schéma était le suivant : un traitement était d'abord administré pendant une certaine période, suivie ensuite d'une période pendant laquelle le patient recevait un placebo ou un traitement alternatif. Ce schéma était alors répété plusieurs fois.

Ce design est utilisé actuellement dans plusieurs domaines comme la psychologie ou encore la médecine (Senn S. , 2019). Les essais *N of 1* sont idéalement utilisés dans l'étude de traitements de longue durée des maladies chroniques, ayant un effet thérapeutique rapide et un effet résiduel négligeable (Duan , Kravitz, & Schmid, 2013).

Comme les deux designs vus précédemment, les essais *N of 1* présentent une série d'avantages mais également quelques inconvénients. Ces avantages et inconvénients vont dépendre de la façon dont les *N of 1* sont mis en place. En effet, comme vu précédemment, ce type d'essais peut être conduit sur un seul patient ou sur plusieurs patients. Dans chacun des cas, les *N of 1* présentent des avantages et des inconvénients.

Dans le cas des essais *N of 1* réalisés sur un seul patient, l'avantage est que ce type d'essai permet d'identifier spécifiquement lequel des traitements testés s'avère être le plus adapté à chaque patient (Duan , Kravitz, & Schmid, 2013). Ces essais sont particulièrement appréciés dans l'étude des maladies rares.

La deuxième possibilité est l'utilisation des essais *N of 1* sur plusieurs individus. Pour rappel, ce type d'études a pour objectif de mettre en évidence un effet traitement et non plus d'étudier l'impact des traitements sur un individu spécifique. Comme dans le cas des essais croisés, le patient est considéré comme son propre témoin. Cela permet d'une part de réduire le nombre d'individus nécessaire à l'étude en conservant un grand nombre d'observations (Duan , Kravitz, & Schmid, 2013). Cela permet d'autre part de différencier la variance intra-patient, puisque les mesures sont répétées chez un même patient, de la variance inter-patient (Senn S. , 2019). De plus, le fait d'administrer de manière répétée plusieurs traitements à un patient permet d'identifier une potentielle interaction patient-traitement (Senn S. , 2019).

Les principaux inconvénients des essais *N of 1* sont présents peu importe que l'essai soit réalisé sur un patient ou sur plusieurs. Un premier inconvénient est lié à la prise de plusieurs traitements successivement. En effet, il existe toujours le risque qu'un traitement ait un effet d'apparition lent ou un effet persistant. Comme vu précédemment, les effets mesurés ne sont alors plus les effets d'un traitement spécifique mais bien de plusieurs traitements en même temps (Duan , Kravitz, & Schmid, 2013). Un deuxième inconvénient est, quant-à-lui, lié aux conditions de l'étude. Étant donné que les essais *N of 1* peuvent être considérés comme de multiples essais croisés, les mêmes conditions sont nécessaires pour pouvoir les mettre en place. Pour rappel, l'objectif étant d'étudier l'effet des traitements sur le long terme, l'état du patient doit rester stable. Il n'est donc pas possible d'utiliser les essais *N of 1* pour étudier des maladies aiguës ou à évolution constante, des traitements à effet permanent ou ayant des

effets lentement réversibles, pouvant être responsables d'un changement d'état du patient (Duan , Kravitz, & Schmid, 2013).

La randomisation en ce qui concerne les essais *N of 1* est un peu différente de celle des essais croisés. En effet, dans les essais *N of 1*, il existe deux types de randomisation (Richard & Naihua, 2019). Le premier type est la randomisation simple. Cette dernière consiste simplement à sélectionner de façon aléatoire lequel des traitements le patient va recevoir à chacune des différentes périodes. Cependant, cette première possibilité comporte plusieurs risques (Richard & Naihua, 2019). En effet, il est possible que le nombre de traitements différents reçus ou que le nombre de périodes d'administration du même traitement ne soient pas équilibrés. Un exemple de cas non équilibré serait une séquence AAAABBAB, dans laquelle le nombre d'administration ainsi que le nombre de séquences successives d'administration du traitement A sont plus importants, comparé au traitement B. La deuxième possibilité est de réaliser des randomisations par bloc, de deux dans cet exemple (AB ou BA) (Richard & Naihua, 2019). Cette deuxième solution permet d'avoir un équilibre entre le nombre d'administrations de chacun des deux traitements et de réduire le nombre maximum de successions du même traitement. Un exemple serait donc ABBABAAB, dans lequel le nombre maximum d'administrations successives du même traitement est 2 et chacun des traitements est administré un même nombre de fois.

A l'heure actuelle, peu d'auteurs se sont intéressés au calcul de la taille d'échantillon des essais *N of 1*. Au vu de la complexité du design, le calcul de la taille d'échantillon dans les essais *N of 1 s*'avère être plus complexe comparé aux essais croisés. En effet, ce calcul dépend de plusieurs facteurs comme par exemple, le nombre de blocs ou la variance intra-patient (Senn S. , 2019). Ce point sera développé par la suite (voir point 4.3).

3 Analyses statistiques

Afin de faciliter la compréhension et la présentation de l'analyse des différents types d'études, la variable Y d'intérêt sera considérée comme étant une variable continue suivant approximativement une distribution normale. Il faut garder à l'esprit que les analyses présentées ici peuvent se généraliser lorsque la variable Y ne suit pas une distribution normale. Les illustrations et les différentes analyses se limitent à l'étude de deux traitements, notés A et B (ou 0 et 1 dans le cas des études *stepped wedge*). Cependant, il est possible dans le cas des essais croisés et des essais *N of 1* de tester plus de deux traitements à la fois.

3.1 Analyses de base des essais croisés

3.1.1 Présentation d'un exemple

Afin d'illustrer au mieux les différentes étapes des analyses statistiques des essais croisés, un exemple de la littérature a été utilisé. Les données de ce dernier proviennent du livre des auteurs Senn et Auclair (Senn & Auclair, 1990). Cette étude a été conçue pour comparer l'effet de deux bronchodilatateurs, le salbutamol et le formoterol, sur l'asthme de l'enfant. La réponse aux traitements est mesurée via le débit expiratoire maximal (noté PEF, de l'anglais, *Peak Expiratory Flow*), une mesure de la capacité pulmonaire. Un total de 14 enfants âgés de 7 à 14 ans atteints d'un asthme, modéré ou sévère, ont été répartis de manière aléatoire dans deux séquences distinctes. Les enfants de la première séquence ont reçu le salbutamol comme premier traitement tandis que les enfants de la seconde ont reçu le formoterol. Ces deux séquences étaient séparées par une période de *wash-out* d'au moins une journée. Les mesures de PEF ont été répétées un certain nombre de fois au cours des deux périodes de traitement. Une première mesure était réalisée 8h après l'administration à la clinique et 3 mesures étaient réalisées par les parents à la maison 10, 11 et 12h après l'administration. Les mesures utilisées pour mettre en évidence une différence éventuelle entre les deux traitements testés sont celles obtenues à la clinique 8h après l'administration. Ces mesures pour chacun des patients sont présentées à la *figure 4*. Ce tableau provient du livre « Cross-over Trials in Clinical Research », écrit par le statisticien Stephen SENN (Senn S. , 2002).

Séquence	N° Patient	PEF (l/min)		
		Formoterol	Salbutamol	Cross-over difference
Formoterol/Salbutamol	1	310	270	40
	4	310	260	50
	6	370	300	70
	7	410	390	20
	10	250	210	40
	11	380	350	30
	14	330	365	-35
Salbutamol/Formoterol	2	385	370	15
	3	400	310	90
	5	410	380	30
	9	320	290	30
	12	340	260	80
	13	220	90	130

Figure 4 : Présentation des données de l'étude du salbutamol et du formoterol, sur l'asthme de l'enfant (Senn & Auclair, 1990). Elles sont présentées par traitement et la 'cross-over difference' est calculée comme la différence entre la mesure de PEF sous formoterol et celle sous salbutamol.

La patient n°8, ayant quitté l'étude après la première mesure de PEF, n'a pas été repris dans le tableau, ni dans les analyses statistiques de ce chapitre. Cet élément illustre bien l'une des difficultés des essais croisés présentés dans l'introduction. En effet, pour ce patient ayant quitté l'étude après l'administration du premier traitement, il manque de l'information. Dans ce cas précis, les scientifiques ont considéré que ne pas tenir compte du patient n°8 ne posait pas de problème pour les analyses (Senn S. , 2002). En effet, ils ont fait l'hypothèse que la raison pour laquelle le patient a quitté l'étude est indépendante de l'étude elle-même donc de l'effet du traitement. L'utilisation des modèles statistiques est une méthode qui permet de contourner les difficultés liées au *'drop out'* des patients. Ce point sera développé plus tard dans l'analyse (voir 3.2). La mesure de PEF est considérée ici comme ayant une distribution normale.

La *figure 4* est composée d'une colonne par traitement ainsi que d'une colonne représentant la différence entre les mesures obtenues après l'administration des deux traitements, appelée 'cross-over difference'. Les différentes valeurs calculées dans cette dernière colonne sont toutes positives, excepté pour le patient n°14.

Dans la littérature, les essais croisés sont présentés le plus souvent par type de traitements (voir *figure 4*). Cependant, il est également possible de présenter les résultats en fonction des différentes périodes d'administration des traitements (voir ci-dessous *figure 5*).

Le tableau suivant provient du livre « Cross-over Trials in Clinical Research », écrit par le statisticien Stephen SENN (Senn S. , 2002) :

Séquence	N° Patient	PEF (l/min)			
		Période 1	Période 2	Différence de périodes	Total
Formoterol/Salbutamol	1	310	270	40	580
	4	310	260	50	570
	6	370	300	70	670
	7	410	390	20	800
	10	250	210	40	460
	11	380	350	30	730
	14	330	365	-35	695
Salbutamol/Formoterol	2	370	385	-15	755
	3	310	400	-90	710
	5	380	410	-30	790
	9	290	320	-30	610
	12	260	340	-80	600
	13	90	220	-130	310

Figure 5 : Présentation des données de l'étude du salbutamol et du formoterol, sur l'asthme de l'enfant (Senn & Auclair, 1990). Elles sont présentées par période et la 'différence de périodes' est calculée comme la différence entre la mesure de PEF de la période 1 et celle de la période 2. La colonne 'Total' est calculée comme étant la somme des mesures des deux périodes pour chacun des patients.

La *figure 5* présente les résultats sous forme d'un tableau composé de différentes colonnes. Les deux premières sont les périodes au cours desquelles les données ont été récoltées. C'est-à-dire, la première période correspond à la mesure de PEF 8h après l'administration du premier traitement et la deuxième période est la mesure réalisée 8h après l'administration du deuxième traitement. La troisième colonne correspond à la soustraction des résultats de la deuxième période aux résultats de la première. Enfin, la dernière colonne correspond à la somme des résultats des deux périodes.

Afin de simplifier la compréhension et amener de la clarté à la suite des analyses un tableau reprenant les notations utilisées par la suite a été réalisé. Le tableau présenté à la *figure 6* a été réalisé sur base de la conception de l'exemple illustré dans ce chapitre. Le tableau

représente donc une étude de deux traitements, notés A et B, réalisée sur deux périodes. La notion Y_{ij} correspond à l'observation de l'individu *i* récoltée à la période *j*. La notation n_1 sera utilisée pour définir le nombre d'individus dans la séquence n°1. La notation Y_i correspond à la différence entre les observations récoltées après l'administration du traitement A et celles récoltées après l'administration du traitement B pour le patient *i*. Enfin, la notation Y_{toti} correspond à la somme des observations effectuées après l'administration des traitements A et B pour le patient *i*.

Séquence		Observations		Cross-over difference	Total
		A	B		
Séquence 1 : A/B	1	Y_{11}	Y_{12}	Y_1	Y_{tot1}
	2	Y_{21}	Y_{22}	Y_2	Y_{tot2}
	⋮	⋮	⋮	⋮	⋮
	n_1	$Y_{n_1 1}$	$Y_{n_1 2}$	Y_{n_1}	Y_{totn_1}
				$\bar{Y}_{seq1}$	$\bar{Y}_{totseq1}$
Séquence 2 : B/A	$n_1 + 1$	$Y_{n_1+1,2}$	$Y_{n_1+1,1}$	Y_{n_1+1}	Y_{totn_1+1}
	$n_1 + 2$	$Y_{n_1+2,2}$	$Y_{n_1+2,1}$	Y_{n_1+2}	Y_{totn_1+2}
	⋮	⋮	⋮	⋮	⋮
	n	$Y_{n,2}$	$Y_{n,1}$	Y_n	Y_{totn}
				$\bar{Y}_{seq2}$	$\bar{Y}_{totseq2}$
				$\bar{Y}_i$	

Figure 6 : Illustration des notations utilisées pour l'analyse des essais croisés

La moyenne de toutes les 'cross over difference' est notée $\bar{Y}_i$. Les moyennes des 'cross over difference' de la séquence 1 et de la séquence 2 sont notées $\bar{Y}_{seq1}$ et $\bar{Y}_{seq2}$, respectivement. Enfin, les moyennes des totaux de la séquence 1 et de la séquence 2 sont notées $\bar{Y}_{totseq1}$ et $\bar{Y}_{totseq2}$, respectivement.

3.1.2 Test t apparié

Pour rappel, les essais croisés permettent d'étudier l'effet individuel de chaque traitement en utilisant le patient comme son propre témoin et de comparer leurs effets. Les 26 observations récoltées dans cette étude peuvent donc être réduites à 13 (2 observations par patient).

La première méthode proposée pour analyser ces paires d'observations est le test t apparié. Cette méthode correspond à l'approche la plus basique. Ce test est réalisé sur des observations dites dépendantes, c'est-à-dire des observations à comparer qui sont liées entre elles. Dans le cas de l'exemple présenté (point 3.1.1), les données qui vont être comparées sont les mesures réalisées après chaque administration des traitements chez un même patient. Étant donné que les observations proviennent du même patient, elles sont considérées comme étant dépendantes. Ce test a comme hypothèse de base que la variable étudiée a une distribution normale, ce qui est supposé dans l'exemple étudié.

Dans cette section, l'objectif et de savoir s'il existe oui ou non une différence d'action entre les deux traitements. Autrement dit : Est-ce qu'un des deux traitements, le salbutamol ou le formoterol, est plus efficace que l'autre ? Ce test est donc utilisé pour étudier l'effet traitement.

L'hypothèse nulle $(H0)$ de ce test est définie comme : l'effet traitement, noté θ, est égal à 0, donc il n'existe pas de différence entre les deux traitements. Par conséquent, l'hypothèse alternative est $(H1)$: l'effet traitement est différent de 0. Elles peuvent se réécrire comme :

$$H0 : \theta = 0$$

$$H1 : \theta \neq 0$$

La statistique de test t s'obtient à l'aide de la formule :

$$t = \frac{\bar{Y}_i}{(\hat{\sigma}_{(1)}/\sqrt{n})} \qquad avec\ t \sim_{H0} Student_{dl=n-1}$$

Dans cette formule, $\bar{Y}_i$ représente l'estimateur de l'effet traitement et la moyenne des différences entre les deux traitements de tous les individus. $\hat{\sigma}_{(1)}/\sqrt{n}$ est l'estimateur de l'écart type de $\bar{Y}_i$ avec n qui représente le nombre de paires et $\hat{\sigma}_{(1)}$ qui est défini comme :

$$\hat{\sigma}_{(1)} = \sqrt{\left\{\sum_{i=1}^{n}(Y_i - \bar{Y}_i)^2/(n-1)\right\}}$$

Le terme Y_i représente les différences obtenues pour chaque individu (*voir figure 6*).

A partir de l'exemple présenté au point 3.1.1, $\bar{Y}_i$ s'obtient en faisant la moyenne des valeurs de la colonne 'cross over difference' dans la *figure 4*. La moyenne générale des 13 différences entre les deux traitements ainsi que l'écart type des différences ont été calculés :

$$\bar{Y}_i = 45.38\ l/min\ et\ \hat{\sigma}_{(1)} = 40.59\ l/min$$

Toutes les informations sont réunies pour calculer la statistique de test :

$$t = \frac{45.38}{(40.59/\sqrt{13})} = \frac{45.38}{11.26} = 4.03$$

La valeur critique de la statistique de test t avec 12 degrés de liberté (dl) et un niveau de significativité α de 5% est de 2.179. Le niveau de significativité, noté α, sera abordé plus en

détails par la suite (voir point 4). L'intervalle de confiance de $\bar{Y}_i$ est calculé comme $\bar{Y}_i \pm t_{0.025}\,\hat{\sigma}_{(1)}/\sqrt{n}$ donc $45.4 \pm 24.5\ l/min$, pour un intervalle de confiance (IC) à 95%. Puisque l'IC à 95% ne reprend pas 0, il existe bien une différence entre la réponse attendue sous formoterol et celle attendue sous salbutamol.

La p valeur est égale à 0.0017. Cette valeur confirme que l'effet traitement est statistiquement significatif, avec un seuil de significativité de 5%.

L'utilisation d'un simple test t apparié suppose que la différence entre les traitements est distribuée de manière aléatoire autour de la vraie différence des traitements. Autrement dit, aucun facteur extérieur n'est considéré comme ayant de l'influence sur la différence entre les traitements. Cependant, il est important de considérer que certains facteurs peuvent impacter cette différence et par conséquent la rendre non totalement aléatoire. Dans ce cas, la différence mesurée n'est plus simplement due aux traitements mais est influencée par un ou plusieurs facteurs extérieurs. Plusieurs exemples vont être décrits afin de bien illustrer ces différents facteurs dans les paragraphes ci-dessous.

Le premier facteur qui pourrait être responsable de ce phénomène est appelé l'*effet période*. Il est possible, par exemple, qu'une tendance générale soit responsable d'un changement d'état du patient qui influencerait les résultats obtenus. Dans ce cas, il est facile d'imaginer que l'administration du même traitement à deux périodes différentes à un patient mènerait à des résultats différents. Ce phénomène a deux conséquences principales.

Dans un premier temps, un effet période a tendance à biaiser l'estimateur de l'effet traitement dans le cas d'un essai non balancé. Un essai non balancé est un essai dans lequel le nombre d'individus dans chaque séquence n'est pas identique. Dans les essais croisés, il est possible que certains individus ne contiennent pas des observations pour chacune des périodes ('drop out'). Ces informations partielles ne peuvent pas être incluses dans les analyses simples présentées ici puisque la variable utilisée pour estimer l'effet traitement est la 'cross over difference'. Dans ce cas, ces individus sont exclus de l'analyse et l'essai est un essai non balancé. Étant donné que dans un essai non balancé la taille des séquences n'est pas équilibrée, chacune des séquences aura un nombre d'observations impactées par l'effet période différent. Dans ce cas, l'effet période va biaiser la moyenne des différences des traitements.

Dans un deuxième temps, cet effet période pourrait avoir un impact sur la variance de l'estimateur de la différence des traitements et sur sa distribution supposée aléatoire. Un exemple simple serait d'imaginer un effet période qui rend les mesures de la deuxième période systématiquement plus élevées, comparé à la première. Dans ce cas précis, la première séquence A/B aura des valeurs systématiquement plus élevées pour les mesures du traitement B tandis que ce sera le traitement A pour la deuxième séquence B/A. Dans le cas d'un essai balancé, les deux séquences auront le même nombre de mesures plus élevées pour la deuxième période. L'estimateur de l'effet traitement ne sera donc pas impacté directement. Bien que l'effet période n'a pas directement d'impact sur cet estimateur, il

pourrait être responsable de l'augmentation de la variance de celui-ci. En effet, les différences pourraient être plus éloignées de la moyenne puisque l'effet période augmentera systématiquement les mesures de la deuxième période. Ainsi la variance de l'estimateur sera augmentée. Dans ce cas, l'hypothèse selon laquelle l'estimation de l'effet traitement est soumise à une variance aléatoire pourrait être erronée. L'effet période pourrait donc rendre cette variance non totalement aléatoire.

Un deuxième facteur, plus complexe, pourrait être l'interaction période-traitement. Ce phénomène apparait lorsque l'effet du traitement est influencé par la période à laquelle celui-ci est administré. Par exemple, si deux évaluations d'un même traitement sont réalisées, une en Janvier et l'autre en Juin, les résultats de celles-ci pourraient être influencés par l'état du patient. En effet, Janvier étant une période plus froide, le patient pourrait présenter, par exemple, de la fièvre à la première visite et pas à la seconde. Si un des traitements est efficace en général, il se peut que son efficacité soit impactée par un changement d'état du patient dû à la période, la fièvre dans ce cas-ci. Ce phénomène s'appelle un effet d'interaction période-traitement. L'effet du traitement peut alors être impacté par cet effet d'interaction période-traitement.

Le troisième facteur est un facteur mentionné précédemment, le *carry over*. Pour rappel, ce phénomène apparait lorsque les effets d'un traitement sont prolongés dans le temps. Les effets mesurés ne sont alors plus les effets des traitements individuels. En prenant l'exemple du point 3.1.1, si un effet *carry over* existait pour le formoterol alors cela mènerait à une différence moins importante pour les patients ayant reçu le formoterol en premier. En observant la *figure 4*, les individus de la séquence formoterol/salbutamol semblent avoir des différences légèrement inférieures, comparés aux individus de la séquence salbutamol/formoterol (Senn S. , 2002). Cependant, cette tendance n'est pas flagrante. Pour réduire ce phénomène, il est possible d'inclure dans l'essai une période dite de *wash-out* comme c'est le cas ici. Pour rappel, cette période correspond à une durée entre les traitements au cours de laquelle les patients ne reçoivent aucun traitement (ou un traitement alternatif). L'effet *carry over*, s'il est présent, dépend directement de la période de *wash-out*. Dans cet essai, la période de *wash-out* est définie comme ayant une durée minimum de une journée.

Un quatrième facteur à mettre en évidence est l'interaction patient-traitement. Un traitement est généralement défini par une action spécifique. Par exemple, l'action spécifique d'un antidouleur est de diminuer voir de supprimer la douleur. Cependant, il est possible que pour certains individus, un traitement ait un effet différent comparé à son action de base. En reprenant l'exemple des antidouleurs, il se peut que pour certains individus un antidouleur fasse moins effet et ne permette pas de diminuer la douleur. Dans l'exemple du point 3.1.1, l'observation n°14 pourrait faire penser à la présence d'un effet d'interaction patient-traitement. En effet dans la *figure 4*, le patient n°14 est le seul patient à présenter une différence négative. Pour l'ensemble des autres patients, la différence entre le formoterol et le salbutamol est positive. Il semblerait donc que, pour le patient n°14, la mesure du

salbutamol soit plus élevée comparée à celle du formoterol. La présence d'un effet d'interaction patient-traitement pourrait donc être une des causes de ce résultat. Toutefois, il est possible que les résultats du patient n°14 soient simplement dus au hasard et reflètent simplement l'instabilité des fonctions pulmonaires chez les asthmatiques. Dans ce cas, cet effet pourrait alors être observé chez n'importe quel autre patient si l'étude était répétée une seconde fois.

L'essai pris comme exemple ne contient que deux périodes. Dans ces conditions, il n'est pas possible de savoir si les mesures obtenues pour le patient n°14 sont dues au hasard ou bien à un effet d'interaction patient-traitement (Senn S. , 2002). Pour faire cette différence, il est nécessaire d'avoir plusieurs prises du même traitement par le patient. Dans ce cas, la différence entre le hasard et l'effet d'interaction aurait pu être faite en fonction des mesures observées. L'idée générale est que si tous les résultats indiquent que le patient n°14 réagit mieux au traitement salbutamol, alors il y aurait bien un effet d'interaction patient-traitement. Tandis que si une différence négative n'avait été observée qu'une seule fois, la conclusion aurait plutôt été que cette différence négative observée est simplement due au hasard.

Un dernier facteur qui pourrait impacter l'effet traitement est l'interaction patient-période. Dans cet exemple, la mesure de PEF peut être impactée par la période à laquelle cette dernière est réalisée. Ce phénomène survient lorsque les patients sont sensibles à certaines variations en fonction de la période. Si le patient est recruté en début d'année (Janvier-Février), il sera plus à risque de tomber malade et avoir des atteintes pulmonaires donc des résultats de PEF impactés, comparé à un patient recruté en milieu d'année (Juillet-Août).

3.1.3 Effet période

Dans la section précédente, les analyses sont basées sur l'hypothèse qu'il n'existe aucune forme d'interaction qui pourrait influencer la différence entre les traitements. Il suffisait alors de regarder la différence entre les effets des deux traitements pour chacun des patients. Mais qu'en est-il s'il existe un effet période pouvant influencer cette différence ? Si un tel effet est présent, alors l'estimateur de l'effet traitement de la section précédente sera biaisé par cet effet période (Senn S. , 2002).

Il existe différentes méthodes qui permettent de tenir compte d'un effet période dans l'estimation de l'effet traitement (Senn S. , 2002). Une première méthode consiste à calculer les différences entre les traitements sur base des deux périodes. L'idée est de soustraire la différence de périodes à la 'cross over difference'. Cette méthode consiste donc à soustraire les moyennes des différences de périodes de chacune des séquences.

Afin de mieux comprendre, cette méthode peut être représentée en prenant le premier patient de chacune des deux séquences.

La différence entre la première période et la deuxième période pour le premier patient ayant reçu la séquence A/B est $Y_{11} - Y_{12}$ (*figure 6*). Cette différence correspond en réalité à la

différence entre le traitement A et le traitement B. Tandis que la différence pour le premier patient de la deuxième séquence B/A est $Y_{n_1+1,1} - Y_{n_1+1,2}$. Cette deuxième différence correspond à la différence entre le traitement B et le traitement A.

Soustraire ces deux différences permet d'éliminer l'effet période :

$$(Y_{11} - Y_{12}) - (Y_{n_1+1,1} - Y_{n_1+1,2})$$

Cette expression correspond en réalité à 2 fois la différence entre le traitement A et le traitement B. Elle doit donc être divisée par deux pour obtenir une estimation de l'effet traitement non biaisée par un effet période. Cette méthode peut être étendue à l'ensemble des données et dans ce cas, ce sont les moyennes des différences de chacune des deux séquences qui sont utilisées. Senn considère donc ici que soustraire les moyennes des différences de périodes, permet de ne conserver que la mesure de l'effet traitement (Senn S. , 2002).

Cette section se concentre, cependant, sur une deuxième méthode. Cette dernière s'adapte plus facilement à des essais plus complexes (Hills & Armitage, 1979). L'objectif ici est de savoir s'il existe un réel effet traitement en tenant compte d'un éventuel effet période. Dans ce cas, un t test peut être réalisé. A la différence de la section précédente (point 3.1.2), l'estimateur de l'effet traitement est obtenu à partir des moyennes des 'cross over difference' de chacune des séquences. Cet estimateur est considéré comme n'étant pas biaisé par un effet période. Ce test a comme hypothèse nulle qu'il n'existe pas de différence entre les deux traitements donc que l'effet traitement, noté θ, est égal à 0. L'hypothèse alternative est donc que l'effet traitement est différent de 0.

$$H0 : \theta = 0$$

$$H1 : \theta \neq 0$$

La statistique de test t est la suivante :

$$t = \frac{\bar{Y}_{seq}}{(\hat{\sigma}_{(2)}/\sqrt{n})} \qquad avec\ t \sim_{H0} Student_{dl=n-2}$$

$\bar{Y}_{seq}$ correspond à l'estimateur de l'effet traitement non baisé par un effet période éventuel. $(\hat{\sigma}_{(2)}/\sqrt{n})$ est l'estimateur de l'écart type de $\bar{Y}_{seq}$.

$\bar{Y}_{seq}$ est obtenu en calculant les moyennes des 'cross-over difference' par séquence et ensuite en calculant la moyenne de ces deux valeurs obtenues. On aura alors, en notant $\bar{Y}_{seq1}$ la moyenne de la séquence 1 et $\bar{Y}_{seq2}$ la moyenne de la séquence 2 :

$$\bar{Y}_{seq} = \frac{\bar{Y}_{seq1} + \bar{Y}_{seq2}}{2} \qquad (1)$$

Afin de simplifier la compréhension de l'analyse, la notation n_2 sera utilisée comme étant le nombre d'individus présents dans la deuxième séquence ainsi $n_2 = n - n_1$. En considérant un essai balancé avec un total de n patients, $n_1 = n_2 = n/2$, la variance de l'estimateur $\bar{Y}_{seq}$ de

l'effet traitement est estimée par $(\hat{\sigma}_{(2)}/\sqrt{n})$. En réalité, il est possible que l'essai ne soit pas balancé (exemple 'drop out'). Pour un essai constitué de n_1 patients dans la première séquence et de n_2 patients dans la deuxième, les variances de chacune des deux moyennes calculées pour les deux séquences sont de ${\hat{\sigma}_{(2)}}^2/n_1$ et ${\hat{\sigma}_{(2)}}^2/n_2$. La variance de l'estimateur $\bar{Y}_{seq}$, devient alors de :

$$\frac{{\hat{\sigma}_{(2)}}^2(1/n_1+1/n_2)}{4} \qquad (2)$$

En assumant que tous les $Y_{1\ldots n}$ ont la même variance inconnue ${\sigma_{(2)}}^2$, alors un estimateur non biaisé de ${\sigma_{(2)}}^2$ est (Hills & Armitage, 1979):

$${\hat{\sigma}_{(2)}}^2 = \frac{(n_1-1)\,S_{seq1}^2 + (n_2-2)\,S_{seq2}^2}{n_1+\,n_2-2}$$

avec :

$$S_{seq1}^2 = \frac{1}{n_1-1}\sum_{i=1}^{n_1}\left(Y_i - \bar{Y}_{seq1}\right)^2$$

$$S_{seq2}^2 = \frac{1}{n_2-1}\sum_{i=n_1+1}^{n}\left(Y_i - \bar{Y}_{seq2}\right)^2$$

Les résultats de (1), sur base de l'exemple étudié et de *la figure 4*, sont alors :

$$\bar{Y}_{seq1} = 30.71\ l/min \text{ et } \bar{Y}_{seq2} = 62.50\ l/min$$

$$\bar{Y}_{seq} = 46.61\ l/min$$

Dans cet exemple, S_{seq1}^2 et S_{seq2}^2 correspondent à 1086.9 (l/min)^2 et 1997.5 (l/min)^2, respectivement et ${\hat{\sigma}_{(2)}}^2$ est alors :

$${\hat{\sigma}_{(2)}}^2 = \frac{(6*1086.9\ \ +5*1997.5)}{6+7-2} = 1500.81\ (\mathrm{l/min})^2$$

En implémentant cette valeur dans (*2*), la variance de l'estimateur $\bar{Y}_{seq}$ est estimée comme :

$$\frac{1500.81(1/7+\ 1/6)}{4} = \ 116.13\ (\mathrm{l/min})^2$$

Tous les éléments sont réunis pour calculer la statistique de test t. Pour rappel, l'hypothèse nulle de ce test est qu'il n'existe pas de différence entre les traitements. La statistique de test t, ayant sous H0 une distribution t, à n_1+n_2-2 degrés de liberté, $(6+7-2)=11$, est :

$$t = \frac{46.61}{\sqrt{116.13}} = \frac{46.61}{10.78} = 4.33$$

Il est simple de constater que les résultats obtenus sont fortement similaires aux résultats obtenus à la section 3.1.2., dans laquelle le résultat de la statistique de test t était de 4.03. La valeur critique de la statistique de test t avec 11 degrés de liberté et un niveau de significativité

α de 5% est de 2.201. Dans l'exemple étudié (point 3.1.1), cette méthode mène donc à la même conclusion que précédemment : il existe bien une différence entre la réponse attendue sous formoterol et celle attendue sous salbutamol.

Afin de savoir laquelle des méthodes utiliser pour estimer et tester l'effet traitement, il est nécessaire de savoir s'il y a, oui ou non, un effet période dans l'exemple analysé. Pour estimer cet effet période, le même principe que celui utilisé pour éliminer l'effet période de l'estimation de l'effet traitement va être appliqué. L'idée ici est de soustraire la 'cross over difference' de la différence de périodes. Pour se faire, il suffit donc de soustraire les moyennes des 'cross over difference' de chacune des séquences afin de ne conserver que l'estimation de la différence des périodes donc de l'effet période. Cette méthode peut être illustrée à partir du premier patient de chacune des séquences de la *figure 6*.

La 'cross over difference' pour le premier patient ayant reçu la séquence A/B est $Y_{11} - Y_{12}$. Cette différence correspond en réalité à la différence entre la Période 1 et la Période 2. La 'cross over difference' pour le premier patient de la deuxième séquence B/A est $Y_{n_1 + 1,2} - Y_{n_1 + 1,1}$. Cette deuxième différence correspond à la différence entre la période 2 et la période 1.

Soustraire ces deux différences permet d'éliminer l'effet traitement et de ne conserver que la différence de périodes :

$$(Y_{11} - Y_{12}) - (Y_{n_1 + 1,2} - Y_{n_1 + 1,1})$$

Cette expression correspond en réalité à 2 fois la différence entre la Période 1 et la Période 2. Elle doit donc être divisée par deux pour obtenir une estimation de l'effet période.

A partir de l'exemple du formoterol et du salbutamol, un test t peut être réalisé. Ce dernier permet de tester s'il y a bel et bien un effet période présent dans l'étude. Ce test a comme hypothèse nulle que l'effet période, noté μ_p, est égal à 0. L'hypothèse alternative est alors que cet effet est différent de 0.

$$H0 : \mu_p = 0$$

$$H1 : \mu_p \neq 0$$

La statistique de test est telle que

$$t = \frac{\hat{P}}{\hat{\sigma}_p^2} \qquad avec\ t \sim_{H0} Student_{dl=n-2}$$

$\hat{P}$ correspond à l'estimation de l'effet période. $\hat{\sigma}_p^2$ représente l'estimateur de la variance de $\hat{P}$.

A partir de la *figure 6*, $\hat{P}$ est estimé comme :

$$\hat{P} = \frac{\overline{Y}_{seq1} - \overline{Y}_{seq2}}{2}$$

Sur base de la *figure 4*, la moyenne des 'cross-over differences' pour la séquence formoterol/salbutamol est de $30.71\ l/min$. Tandis que la moyenne de la séquence salbutamol/formoterol est de $62.5\ l/min$.

L'estimation de l'effet période correspond donc à $(30.71 - 62.5)/2 = -15.90$ l/min. L'erreur standard reste la même que celle calculée pour la différence des effets des traitements en tenant compte d'un effet période. En effet, la variance de $\hat{P}$ correspond à :

$$\hat{\sigma}_p^2 = \frac{\hat{\sigma}_{(2)}{}^2(1/n_1 + 1/n_2)}{4}$$

Donc, elle est égale à $116.13\ (l/min)^2$. L'écart type de cet estimateur est donc de $10.776\ l/min$. La statistique de test t ayant sous H0 une distribution t, à 11 degrés de liberté, dont l'hypothèse nulle est qu'il n'y a pas d'effet période, correspond alors à $-15.90/10.78 = -1.47$. Le résultat n'étant pas significatif, l'hypothèse nulle ne peut pas être rejetée. La conclusion est qu'il n'y pas assez de preuves pour mettre en évidence un effet période dans cet essai.

Puisque le test est non significatif, il est préférable d'estimer l'effet traitement via l'approche détaillée à la section 3.1.2. Si le résultat avait été significatif et qu'un effet période avait été mis en évidence, il aurait été préférable dans ce cas d'utiliser l'approche détaillée à la section 3.1.3.

3.1.4 Effet *carry over* et interaction traitement-période

L'effet *carry over* et l'interaction traitement-période sont deux facteurs qui pourraient influencer l'estimation de l'effet traitement. Il est donc important d'être conscient de la présence ou non de ces effets dans l'essai. En réalité dans un essai croisé AB/BA (deux traitements/deux périodes), ces deux effets ne peuvent pas être estimés séparément. Cela signifie que si la présence d'un effet *carry over* est détectée, il faut considérer qu'il y a également présence d'un effet d'interaction traitement-période.

Pour tester la présence de cet effet, il n'est pas possible d'utiliser les données de la 'cross-over différence' pour deux raisons (Senn S. , 2002). La première raison est qu'une différence entre les 'cross-over difference' à l'intérieur d'une séquence peut avoir trois explications : la variation aléatoire, l'interaction patient-traitement et l'interaction patient-période. Donc connaitre les différences qui existent à l'intérieur d'une même séquence n'apporte aucune information concernant le *carry over*. Deuxièmement, les moyennes des différences des effets des patients et les moyennes des différences par séquence ont été utilisées précédemment pour mesurer l'effet traitement et l'effet période, respectivement. L'hypothèse lors de l'estimation de ces effets était qu'aucun effet *carry over* n'était présent mais que ce soit le cas ou non, ces moyennes mesureront en partie l'effet traitement ou l'effet période. Ces données ne permettent donc pas d'obtenir une estimation spécifique d'un effet *carry over*.

Pour pouvoir étudier l'effet *carry over*, Senn propose d'utiliser le total des réponses aux traitements des deux périodes pour chaque patient (Senn S. , 2002). Pour se faire, il est

nécessaire de considérer que la différence entre les totaux de deux patients de deux séquences différentes ne peut pas venir d'un effet traitement puisque tous les patients ont reçu les deux traitements, ni d'un effet période puisque chaque patient a été suivi pendant les deux périodes. Si l'effet traitement persiste au cours de la deuxième période alors un patient ayant reçu la séquence de traitements A/B aura un effet *carry over* provenant de A et un patient ayant reçu B/A aura un *carry over* provenant de B. La mise en évidence d'une différence des totaux entre deux patients de deux séquences différentes reflète alors la présence d'un effet *carry over*. Cette méthode restera efficace tant que les deux traitements ont des effets *carry over* différents.

Pour investiguer la présence ou non d'un effet *carry over*, un test t peut être réalisé. Ce test a comme hypothèse nulle que la différence entre les moyennes des totaux des deux séquences, notée μ_t, est égale à zéro et donc qu'il n'y a pas d'effet *carry over*. L'hypothèse alternative est donc que cette différence n'est pas égale à 0.

$$H0 : \mu_t = 0$$

$$H1 : \mu_t \neq 0$$

La statistique de test est :

$$t = \frac{Y_{totdiff}}{(2\hat{\sigma}_{(3)}/\sqrt{n})} \qquad avec\ t \sim_{H0} Student_{dl=n-2}$$

$Y_{totdiff}$ représente l'estimateur de la différence entre les moyennes des totaux des deux séquences. $2\hat{\sigma}_{(3)}/\sqrt{n}$ est l'estimateur de l'écart type de $Y_{totdiff}$ dans le cas où l'essai est balancé et donc $n_1 = n_2 = n/2$.

L'estimateur de la différence entre les deux moyennes des totaux de chaque séquence, avec $\bar{Y}_{totseq1} = Moyenne\ totaux\ séquence\ 1$ et $\bar{Y}_{totseq2} = Moyenne\ totaux\ séquence\ 2$, est obtenu comme suit :

$$Y_{totdiff} = \bar{Y}_{totseq1} - \bar{Y}_{totseq2}$$

Comme vu précédemment, il est possible que l'essai ne soit pas balancé. Dans ce cas, la variance de l'estimateur $Y_{totdiff}$ est :

$$\hat{\sigma}_{(3)}{}^2(1/n_1 + 1/n_2)$$

En assumant que tous les $Y_{tot1\ldots n}$ ont la même variance inconnue $\sigma_{(3)}{}^2$ alors un estimateur non biaisé de $\sigma_{(3)}{}^2$ est :

$$\hat{\sigma}_{(3)}{}^2 = \frac{(n_1 - 1)\, S_{tot1}^2 + (n_2 - 2)\, S_{tot2}^2}{n_1 + n_2 - 2}$$

Avec :

$$S_{tot1}^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}\left(Y_{toti} - \bar{Y}_{totseq1}\right)^2$$

$$S_{tot2}^2 = \frac{1}{n_2 - 1} \sum_{i=n_1+1}^{n} \left(Y_{toti} - \bar{Y}_{totseq2}\right)^2$$

Le test t peut être réalisé sur base de l'étude du salbutamol et du formoterol. Pour réaliser ce test, il faut procéder comme précédemment mais en utilisant les données de la colonne 'Total', qui correspondent aux totaux, de la *figure 5*. La différence des moyennes est estimée comme suit :

$$\bar{Y}_{totseq1} = 643.57\ l/min \text{ et } \bar{Y}_{totseq2} = 629.17\ l/min$$

$$Y_{totdiff} = 14.40\ \text{l/min}$$

Avec $S_{tot1}^2 = 13\,072.62$ et $S_{tot2}^2 = 30\,264.17$, l'estimateur $\hat{\sigma}_{(3)}{}^2$ est égal à :

$$\hat{\sigma}_{(3)}{}^2 = \frac{(6 * 13072.62 \ + 5 * 30264.17)}{6 + 7 - 2} = 20\,886.96\ (\text{l/min})^2$$

L'estimation de la variance de $Y_{totdiff}$ est alors :

$$20886.96(1/6 + 1/7) = 6\,465.01(\text{l/min})^2$$

L'écart type de l'estimateur est donc égal à $\sqrt{6465.01} = 80.41\ l/min$.

La statistique de test ayant une distribution t de Student sous H_0 est $\frac{14.40}{80.41} = 0.18$.

La valeur obtenue de la statistique de test, avec 11 degrés de liberté, conduit au non rejet de l'hypothèse nulle. Il n'y pas assez de preuves pour affirmer que la différence entre les totaux des deux séquences est différente de 0.

Il est possible d'estimer un effet *carry over* à partir de l'estimation de la différence des totaux obtenue. Afin de comprendre comment obtenir cette estimation, un exemple simple d'un essai croisé va être illustré à partir du premier patient de chacune des séquences (*figure 6*).

En considérant un essai croisé AB/BA à deux périodes, un effet *carry over* peut être présent pour le traitement A, noté γ_A ou pour le traitement B, noté γ_B. L'effet traitement est estimé à partir de la différence entre A et B de chacune des deux séquences (voir 3.1.3). Les mesures récoltées pour les patients de la séquence A/B, et en particulier pour cette illustration le premier patient $(i = 1)$, pourraient être impactées par un effet *carry over* du traitement A. Dans le cas où un tel effet est présent, alors l'observation réalisée après l'administration du traitement A est Y_{11} tandis que l'observation réalisée après l'administration du traitement B correspond en réalité à $Y_{12} + \gamma_A$. La 'cross over difference' de ce patient sera égale à $Y_{11} - Y_{12} - \gamma_A = Y_1 - \gamma_A$. Tandis que les mesures récoltées pour les patients de la séquences B/A, et en particulier pour cette illustration le premier patient de cette séquence $(i = n_1 + 1)$, pourraient être impactées par un effet *carry over* du traitement B. La première observation réalisée sur ce patient après l'administration du traitement B correspond à $Y_{n_1+1,1}$ tandis que la deuxième observation réalisée après l'administration du traitement A correspond à $Y_{n_1+1,2} + \gamma_B$. Alors la 'cross over difference' pour ce patient est en réalité : $Y_{n_1+1,2} + \gamma_B - Y_{n_1+1,1} = Y_{n_1+1} + \gamma_B$.

Dans ce cas, la moyenne des 'cross over difference' de ces deux patients provenant chacun d'une séquence différente sera égale à :

$$\frac{(Y_1 - \gamma_A) + (Y_{n_1+1} + \gamma_B)}{2} = \frac{(Y_1 + Y_{n_1+1})}{2} + \frac{(\gamma_B - \gamma_A)}{2}$$

Si on considère maintenant l'ensemble des patients et non plus les deux premiers des deux séquences alors la moyenne des 'cross over difference' de tous les patients correspond à :

$$\frac{(\bar{Y}_{seq1} + \bar{Y}_{seq2})}{2} + \frac{(\gamma_B - \gamma_A)}{2}$$

En présence d'un effet *carry over*, l'estimation de l'effet traitement ne correspond plus à la moyenne des moyennes des deux séquences mais est biaisée par le terme $(\gamma_B - \gamma_A)/2$ $(\text{Senn S.}, 2002)$. Ce terme correspond en réalité à la différence des effets *carry over* des deux traitements. Ce dernier peut être estimé à partir de la différence des moyennes des totaux de chaque séquence. Il suffit pour cela d'inverser le signe de la différence des totaux et de la diviser par deux puisqu'en réalité la différence des totaux correspond à :

$$(\bar{Y}_{totseq1} + \gamma_A) - (\bar{Y}_{totseq2} + \gamma_B)$$

L'estimation de $(\gamma_B - \gamma_A)/2$ pour l'exemple du formoterol et du salbutamol sera égal à $-(14.40/2) = -7.2\ l/min$. A savoir que lorsque les effets *carry over* du traitement A et du traitement B sont nulles donc $\gamma_A = \gamma_B = 0$ ou que ces effets sont parfaitement identiques, $\gamma_A = \gamma_B$ alors $(\gamma_B - \gamma_A)/2$ sera égal à 0.

Bien qu'une estimation d'un potentiel "effet *carry over*" ait été obtenue, il faut garder à l'esprit que pour appliquer cette méthode l'auteur a émis l'hypothèse que l'effet période ainsi que l'effet traitement sont les mêmes pour tous les individus. L'effet carry over, s'il est présent, est également considéré comme étant identique pour tous les patients. De plus cette méthode ne tient pas compte du fait qu'il pourrait exister un effet patient. Si ce dernier existe, alors le total des traitements sera impacté par cet effet patient et la différence des totaux ne mesurera plus uniquement l'effet *carry over*. Avec cette méthode, il n'est donc pas possible de confirmer avec certitude qu'il n'existe réellement pas d'effet *carry-over* au vu du nombre de facteurs qui pourraient influencer le total utilisé pour estimer cet effet. Une solution proposée est l'utilisation de modèles statistiques. En effet, les modèles permettraient de tenir compte de cet effet patient afin de rendre la mesure de l'effet *carry over* plus précise. Cette méthode sera détaillée par la suite (voir point 3.2.3.3).

3.2 Modélisation statistique des essais croisés

La présentation de l'analyse des essais croisés dans cette section est principalement basée sur la troisième édition du livre intitulé '*Applied Mixed Models in Medicine*' (Brown & Prescott, 2015).

Les analyses, décrites dans la partie 3.1, permettent d'étudier l'effet traitement de manière simple. Elles permettent également de prendre en compte un éventuel effet période et d'enquêter sur la présence ou non d'un effet *carry over*, effet difficile à estimer comme décrit précédemment. Au vu de la complexité des données récoltées dans les essais croisés, il est préférable d'utiliser des modèles statistiques. En effet, comme expliqué précédemment, il est courant dans les essais croisés que des patients n'aient que certaines observations utilisables, par exemple si le patient quitte l'étude tôt ou ne se présente pas à certaines visites ('drop out'). Dans ce cas, il est nécessaire d'utiliser des modèles statistiques pour tenir compte de ces informations partielles, ce qui n'est pas possible avec des analyses statistiques plus simples. De plus, les modèles statistiques plus complexes, comme les modèles mixtes, permettent de prendre en compte le fait que les mesures sont répétées au cours du temps ou encore qu'elles ont différentes origines (différents centres, hôpitaux, …).

Plusieurs modèles différents, du plus simple au plus complexe, vont être décrits pour bien comprendre l'utilité et les avantages de chacun d'entre eux. Pour simplifier la compréhension et la représentation des modèles, les effets fixes seront représentés par la lettre $\boldsymbol{\alpha}$, les effets/coefficients aléatoires par la lettre $\boldsymbol{\beta}$ et les termes d'erreur des modèles par la lettre $\boldsymbol{\epsilon}$. Il est considéré ici qu'une seule mesure est réalisée après l'administration de chaque traitement.

3.2.1 Visualisation de la base de données de l'exemple

La base de données de l'exemple sur l'étude du salbutamol et du formoterol, sur l'asthme de l'enfant a été reproduite (Senn & Auclair, 1990).

Afin de mieux visualiser les données et faciliter la compréhension des analyses statistiques, les données des 5 premiers patients sont représentées dans le tableau ci-dessous :

ID	PEF	Treat	Period	Carry
1	270	0	1	1
1	310	1	0	1
2	370	0	0	0
2	385	1	1	0
3	310	0	0	0
3	400	1	1	0
4	260	0	1	1
4	310	1	0	1
5	380	0	0	0
5	410	1	1	0

Figure 7 : Représentation des données des 5 premiers individus de la base de données de l'étude du salbutamol et du formoterol, sur l'asthme de l'enfant.

Dans le tableau ci-dessus, chacun des patients est représenté par deux lignes puisque dans cette étude deux observations par patient ont été récoltées. La colonne **ID** représente le numéro attribué au patient. La colonne **PEF** correspond aux mesures du débit expiratoire maximal, réalisées deux fois pour chacun des patients. La colonne **Treat** correspond au traitement administré avant la mesure, avec le salbutamol comme traitement de référence. La colonne **Period** représente la période à laquelle la mesure a été réalisée. Pour rappel, dans cette étude, deux mesures ont été réalisées, une après chaque administration de traitements, deux périodes sont donc étudiées. La période n°1 est la période prise comme référence. La dernière colonne, nommée **Carry**, correspond à la séquence des traitements reçus par le patient. La séquence prise comme référence est la séquence salbutamol/formoterol.

Le patient n°1 (ID = 1) peut être pris comme exemple pour illustrer la lecture de ce tableau. La séquence formoterol/salbutamol a été attribuée de façon aléatoire à ce patient. Suite à l'administration du formoterol en première période, le patient n°1 a eu une mesure de PEF de 310 l/min. Ensuite, une deuxième mesure a été réalisée après l'administration du salbutamol à la deuxième période. Le PEF de cette deuxième mesure était de 270 l/min.

3.2.2 Modèles à effets fixes

3.2.2.1 Effet traitement

Le premier modèle envisagé pour étudier l'effet traitement est un modèle simple, avec peu de paramètres.

L'observation y_{ij} du patient i ayant reçu le traitement j en considérant que A est le traitement de référence correspond à :

$$y_{ij} = \mu + \alpha_{1,j} treat_j + \epsilon_{ij} \qquad i = 1, \dots, n \quad j = A, B \qquad (3)$$

avec

μ : terme d'intercept

$\alpha_{1,j}$: effet traitement

ϵ_{ij} : résidus du patient *i* ayant reçu le traitement *j*

A est considéré comme le traitement de référence et $treat_j$ prendra la valeur de 0 si l'observation est réalisée suite à la prise du traitement *A* et 1 si B. Donc $\alpha_{1,j}$ représente l'effet du traitement B par rapport au traitement A. Le terme d'intercept correspond à la valeur moyenne de la variable réponse avec le traitement *A.*

Le terme d'erreur, ϵ_{ij}, correspond à $\epsilon_{ij} = y_{ij} - \mu - \alpha_{1,j} treat_j$ pour chacun des patients à chacune des périodes. Ils sont considérés comme indépendants, identiquement distribués et ayant une distribution normale :

$$\epsilon_{ij} \sim N(0, \sigma^2)$$

Cette hypothèse peut être simplement vérifiée, par exemple, à l'aide d'un graphique (histogramme ou QQplot) des résidus, notés e_{ij}. Les résidus représentent la différence entre la valeur observée et la valeur prédite par le modèle.

Étant donné que μ et $\alpha_{1,j}$ sont deux paramètres constants, la variance des observations individuelles y_{ij} est égale à la variance des résidus, σ^2. Dans ce modèle, la covariance entre deux observations est nulle puisque les observations sont considérées comme indépendantes les unes des autres.

$$Var(y_{ij}) = \sigma^2 \text{ et } Cov(y_{ij}, y_{i'j'}) = 0$$

La variance des résidus, σ^2, peut être estimée à l'aide de la méthode ANOVA, 'Analysis of Variance' (Kutner, Nachtsheim, Neter, & Li, 2004). Cette méthode de base a pour principe de décomposer la variance totale des données afin d'en expliquer la cause. Dans l'étude sur les effets du salbutamol et du formoterol, par exemple, une partie de la variance pourrait être expliquée par les traitements. En effet, le formoterol pourrait être associé à des valeurs élevées de PEF tandis que le salbutamol à des valeurs plus faibles. En réalité pour l'exemple du formoterol et du salbutamol, ce premier modèle, composé d'une réponse quantitative (PEF) et d'une variable explicative qualitative, peut être estimé par la méthode ANOVA. Le terme d'intercept correspond à la moyenne des valeurs de PEF obtenues suite à l'administration du traitement salbutamol. L'effet traitement $\alpha_{1,j}$ correspond à la différence entre la moyenne des valeurs obtenues suite à l'administration du salbutamol et la moyenne des valeurs obtenues suite à l'administration du formoterol.

3.2.2.2 Effet patient

Jusqu'à présent, le fait que les observations sont récoltées plusieurs fois chez un même patient et donc pairées n'a pas été pris en compte. Pour se faire, un effet patient est ajouté au premier modèle à effets fixes (*3*). Le modèle devient alors :

$$y_{ij} = \mu + \alpha_{1,j} treat_j + \alpha_{2,i}{}' x_i + \epsilon_{ij} \qquad i = 1, \dots, n \quad j = A, B \qquad (4)$$

avec

$\alpha_{2,i}{}'$: vecteur des effets patient

$\alpha_{2,i}{}'$ représente le vecteur des effets patient. Étant donné que ces paramètres sont considérés comme fixes, une estimation sera obtenue pour chacun des patients présents dans l'étude. Le nombre de paramètres estimés est de $n-1$ puisqu'un des patients est considéré comme la référence. L'estimation de l'effet de ce patient de référence est incluse dans le terme d'intercept. Le terme $\alpha_{2,i}{}' x_i$ se décompose comme :

$$\alpha_{2,i}{}' x_i = \alpha_{2,i=1} x_1 + \alpha_{2,i=2} x_2 + \cdots + \alpha_{2,i=n-1} x_{n-1}$$

$\hat{\alpha}_{2,i}$ est à l'estimateur de l'effet du patient *i* ; x_i prendra la valeur de 1 lorsque l'observation est réalisée sur le patient *i*, avec $i = 1, \dots, n-1$. L'estimation de l'effet du dernier patient, le patient n, sera incluse dans le terme d'intercept.

Dans ce modèle, la variance des résidus σ^2 correspond à la variabilité à l'intérieur d'un même patient. Elle devrait être plus faible que dans le modèle précédent puisque le fait que les mesures sont répétées par patient a été pris en compte.

En reprenant l'exemple illustré au point 3.1.1, il est considéré ici que le traitement de référence est le salbutamol. L'estimateur, $\hat{\mu}$, correspond donc à l'effet moyen du salbutamol $(j = A)$ tandis que $\hat{\mu} + \hat{\alpha}_{1,j} treat_j$ correspond à l'estimation de l'effet moyen du formoterol $(j = B)$. Dans ce cas, l'estimateur $\hat{\alpha}_{1,j}$ correspond en réalité à l'estimation de la différence qui existe entre les deux traitements. L'application de ce modèle sur les données du formoterol et du salbutamol permet d'obtenir une estimation de l'effet traitement, $\hat{\alpha}_{1,j}$. Un test t est réalisé suite à cette estimation. En réalité, le test réalisé sur les paramètres fixes d'un modèle est appelé un test de Wald. Le test de Wald permet de tester des hypothèses, comme par exemple l'égalité de plusieurs paramètres, sur les effets fixes. Lorsque le test de Wald est réalisé sur un seul paramètre, la distribution asymptotique de la statistique de test est une normale (Wasserman, 2004). Étant donné que le test t peut être approximé par une normale lorsque *n* est grand alors le test t et le test de Wald mèneront aux mêmes conclusions (rejet ou non de H_0), dans le cas d'un test réalisé sur un paramètre fixe unique (Wasserman, 2004).

Ce test a comme hypothèse nulle que le paramètre $\alpha_{1,j}$ est égal à 0 et donc que l'effet traitement n'est pas significativement différent de zéro. Son hypothèse alternative est donc que le paramètre $\alpha_{1,j}$ est différent de 0.

$$H0 : \alpha_{1,j} = 0$$

$$H1 : \alpha_{1,j} \neq 0$$

L'estimation de l'effet traitement du modèle est de $45.38\ l/min$ et la statistique de test t est égale à 4.03 avec une p valeur de 0.0017. Puisque la p valeur est significative avec un seuil de significativité de 5%, l'hypothèse nulle est rejetée. Les résultats obtenus sont donc équivalents aux résultats obtenus au point 3.1.2.

3.2.2.3 Effet période

Comme vu précédemment, il serait intéressant de prendre en compte dans le modèle l'effet période. Pour rappel, les essais croisés sont des études réalisées sur plusieurs périodes. Il est donc possible qu'une tendance générale soit responsable d'un changement d'état du patient qui influencerait les résultats obtenus donc l'effet des traitements étudiés. Pour prendre en compte cet effet, il suffit de rajouter au modèle précédent (*4*) un effet période. Le modèle pour l'observation du patient *i* à la période *k* ayant reçu le traitement *j* devient alors :

$$y_{ijk} = \mu + \alpha_{1,j} treat_j + \alpha_{2,i}{}' x_i + \alpha_{3,k} p_k + \epsilon_{ijk} \quad (5)$$

$$i = 1, \dots, n \quad j = A, B \quad k = 1,2$$

avec

$\alpha_{3,k}$: effet période

Quelques petites précisions sont à faire par rapport à cet effet. Dans ce modèle, la variable période peut être incluse de deux manières : soit comme une variable qualitative, soit comme une variable quantitative. Dans le cas où la période est considérée comme une variable qualitative, un effet période est estimé pour chacune des périodes de l'étude (avec une période de référence). Étant donné que la présentation des techniques d'analyse des essais croisés est basée sur le cas de deux traitements, il y a donc deux périodes de mesures. L'estimation de la période de référence sera incluse dans le terme d'intercept et le paramètre $\alpha_{3,k}$ représentera alors la différence entre les deux périodes. Dans le cas où la variable est considérée comme quantitative, $\alpha_{3,k}$ correspond à l'effet période et p_k correspond à la période (période n°1 alors $p_k = 1$, période n°2 alors $p_k = 2$). Dans ce deuxième contexte, l'intercept correspond à l'estimation de la variable réponse lorsque p_k vaut 0. Généralement dans les essais croisés AB/BA, la période est considérée comme une variable qualitative.

Pour reprendre l'exemple de l'étude du formoterol et du salbutamol, l'estimation de l'effet traitement est de $46.61\ l/min$. Suite à cette estimation, un test t est réalisé. Ce test a comme hypothèse nulle que le paramètre de l'effet traitement, $\alpha_{1,j}$, en tenant compte d'un potentiel effet période est égal à 0, donc qu'il n'y a pas de différence entre les traitements. L'hypothèse alternative est que ce paramètre est différent de 0.

$$H0 : \alpha_{1,j} = 0$$

$$H1 : \alpha_{1,j} \neq 0$$

La statistique du test de Wald est égale à 4.32 avec une p valeur de 0.0012. Cette valeur confirme que l'effet traitement en prenant en compte la présence d'un potentiel effet période est statistiquement significatif, avec un seuil de significativité de 5%. Ce résultat est donc similaire à celui calculé au point 3.1.3.

Dans cet exemple, la variable période a été considérée comme étant une variable qualitative. L'estimation de l'effet de la période n°1 $(k = 1)$, prise comme période de référence, est incluse dans le terme d'intercept. Le paramètre $\alpha_{3,k=2}$ correspond à la différence entre la période n°1 et la période n°2. L'estimation de ce paramètre est de $-15.89\ l/min$ et donc semblable à celle obtenue au point 3.1.33.1.3.

Un test de Wald est réalisé sur ce paramètre $\alpha_{3,2}$. Ce dernier a comme hypothèse nulle que le paramètre $\alpha_{3,2}$ est égal à 0, donc qu'il n'y pas de différence entre les deux périodes. L'hypothèse alternative est, quant-à-elle, que ce paramètre est différent de 0.

$$H0 : \alpha_{3,2} = 0$$

$$H1 : \alpha_{3,2} \neq 0$$

La valeur de la statistique de test est de -1.47. L'hypothèse nulle ne peut pas être rejetée. Le modèle confirme bien que dans l'étude du formoterol et du salbutamol, l'effet période n'est pas significatif.

3.2.2.4 Mesure de référence

Dans les essais croisés, il se peut que des mesures soient réalisées avant toute administration de traitements. Ces observations correspondent donc à l'état initial du patient, un état de référence. Dans ce cas, ces mesures sont réalisées au moment de la randomisation avant toute administration des différents traitements ('baseline'). Il est raisonnable de supposer qu'il existe une relation entre la mesure pré-traitement ('baseline') et les mesures post-traitement réalisées chez un même patient. Pour étudier cette relation, il est nécessaire de récolter et donc d'avoir à disposition des données avant toute administration des traitements. Si une mesure initiale a été réalisée chez chacun des patients alors le modèle de base (*3*) devient (Brown & Prescott, 2015) :

$$y_{ij} = \mu + \alpha_{1,j} treat_j + \alpha_{2,b} b_i + \epsilon_{ij} \qquad i = 1, \dots, n \quad j = A, B$$

avec

$\alpha_{2,b}$: effet de la mesure initiale/ de référence

$\alpha_{2,b}$ représente un terme constant estimé grâce aux données et b_i, une covariable du modèle qui correspond à la valeur initiale pour chacun des patients. L'utilisation d'une mesure de référence va permettre d'augmenter la précision de l'estimation de l'effet traitement (Brown & Prescott, 2015).

Dans l'exemple de l'étude du salbutamol et du formoterol, certains patients pourraient avoir naturellement une valeur de PEF plus élevée. Dans ce cas, il serait alors raisonnable d'utiliser

une mesure initiale. Celle-ci permettrait d'augmenter indirectement la précision des mesures réalisées pour chacun des patients (Senn S. , 2002).

Le cas le plus simple des mesures de référence a été présenté ici, c'est-à-dire avant toute administration de traitements en début d'étude. Cependant, il existe de nombreux articles dans lesquels les auteurs discutent de l'utilisation et des différentes manières de récolter ces valeurs de référence. Senn fait référence à plusieurs types de mesures (Senn S. , 2002). Ces différents types de baseline sont l'unique mesure récoltée avant l'administration du premier traitement, celles récoltées avant chaque administration, celle prise à la fin du premier traitement avant le commencement du deuxième traitement et celle récoltée à la fin du deuxième traitement. Bien que s'il existe un effet *carry over*, la mesure réalisée avant l'administration de tout traitement est en réalité le seul type qui soit une mesure de référence (Senn S. , 2002). Dans ce contexte, afin de simplifier la présentation des différentes analyses des essais croisés, seulement le premier type est présenté dans ce travail.

3.2.3 Modèles mixtes

Comme expliqué précédemment, l'utilisation de modèles mixtes va permettre, soit via des modèles à covariance pattern soit via des effets aléatoires, de prendre en compte le fait que les données ne sont pas indépendantes. L'estimation des paramètres des modèles mixtes s'avère être plus complexe, comparée à celle des modèles linéaires simples. Plusieurs méthodes existent comme la méthode du maximum de vraisemblance ou encore la méthode du maximum de vraisemblance restreinte (Molenberghs & Verbeke, 2009) (Fitzmaurice, Laird, & Ware, 2004). Ces méthodes ne seront pas développées dans ce travail.

Deux types de modèles mixtes vont être présentés. Le premier est un modèle qu'on appelle modèle à 'Covariance pattern'. Ce premier modèle permet de tenir compte de la corrélation entre les observations, venant d'un même individu, grâce à la matrice **R**. Cette matrice correspond à la matrice de covariance des résidus du modèle. Le second type de modèles sont des modèles appelés des modèles à effets aléatoires. Ce type de modèles permet de tenir compte de la corrélation entre les données directement à partir de ces effets aléatoires.

3.2.3.1 Effet patient

Dans les modèles précédents, l'ensemble des termes des différents modèles étaient considérés comme fixes et donc sans distribution propre. Une approche différente serait de considérer certains des termes comme étant des réalisations d'une distribution de probabilité.

Sur base du modèle à effet patient fixe (*4*), le même modèle peut être créé mais avec l'effet patient qui est considéré, cette fois, comme un effet aléatoire. Dans ce cas, le modèle est :

$$y_{ij} = \mu + \alpha_{1,j} treat_j + \beta_{1,i} + \epsilon_{ij} \qquad i = 1, \dots, n \quad j = A, B \qquad (6)$$

avec

$\alpha_{1,j}$: effet traitement

$\beta_{1,i}$: effet du patient *i*

L'intercept μ correspond à la moyenne de tous les patients pour le traitement de référence. Dans ce modèle, l'effet patient $\beta_{1,i}$ n'est plus considéré comme fixe mais bien comme un effet aléatoire ayant une distribution normale.

$$\beta_{1,i} \sim N(0, {\sigma_x}^2)$$

Pour chacun des patients, un $\hat{\beta}_{1,i}$ peut être prédit et il correspond à la déviation de l'intercept du patient *i* par rapport à la moyenne de tous les patients. Ils ne sont plus considérés comme des paramètres ici mais comme des réalisations d'une distribution aléatoire. Ils sont généralement supposés indépendants et identiquement distribués. Ce modèle est appelé un modèle à effet aléatoire.

Le modèle présenté peut se réécrire sous une forme matricielle : $Y = X\alpha + Z\beta + \epsilon$, avec X et Z les matrices designs des effets fixes et aléatoires, respectivement.

A partir des 2 premiers patients de l'exemple du formotérol et du salbutamol (*figure 7*) avec $X_{i,1:2}$ et $Z_{i,1:2}$ les matrices designs des effets fixes et aléatoires des deux premiers patients, respectivement, et en considérant le salbutamol comme traitement de référence, la forme matricielle est :

$$X_{i,1:2} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 1 & 0 \\ 1 & 1 \end{pmatrix} \quad \alpha = (\mu, \alpha_{1,B})' \quad \textbf{et} \quad Z_{i,1:2} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \quad \beta = (\beta_{1,1}, \beta_{1,2})'$$

Les paramètres d'intérêt de ce modèle sont donc $\alpha_{1,j}$ et ${\sigma_x}^2$. A noter qu'ajouter l'effet patient aléatoire au modèle ne rajoute qu'un seul paramètre et ce peu importe le nombre de patients présents dans l'étude.

Les termes d'erreur sont supposés indépendants, identiquement distribués ayant une distribution normale comme précédemment :

$$\epsilon_{ij} \sim N(0, \sigma^2)$$

$\beta_{1,i}$ et ϵ_{ij} sont supposés indépendants.

Chacun des termes aléatoires du modèle donnera lieu à une composante de variance. En ce qui concerne l'effet patient, cette composante est notée ${\sigma_x}^2$ et représente la variation qui existe entre les patients, appelée la 'variabilité inter-patient'. La variance des résidus σ^2 représente toujours, quant-à-elle, la variabilité qui existe entre les observations récoltées chez un même patient, appelée 'variabilité intra-patient'. La variance des observations individuelles y_{ij} dans ce modèle à effets aléatoires se calcule comme la somme des toutes les composantes de variance et est égale à :

$$Var(y_{ij}) = {\sigma_x}^2 + \sigma^2$$

Dans ce modèle, les observations provenant de deux patients différents sont toujours considérées comme indépendantes. Cependant, les observations d'un même patient (i = i') ne sont pas considérées comme indépendantes puisque :

$$Cov(y_{ij}, y_{i'j'}) = \sigma_x{}^2 \qquad \text{pour } i = i'$$

Donc

$$Cor(y_{ij}, y_{i'j'}) = \frac{\sigma_x{}^2}{\sigma_x{}^2 + \sigma^2} \qquad \text{pour } i = i'$$

3.2.3.2 Effet période

Il est possible de tenir compte de la corrélation entre les observations répétées chez un même patient avec un autre type de modèles. Ces modèles sont les modèles à covariance pattern.

Il faut cependant tenir compte du fait que ce type de modèles sont des modèles qui sont plus adaptés à des essais qui contiennent de nombreuses périodes. Ils sont donc généralement utilisés dans des essais croisés contenant plus de périodes que les essais croisés AB/BA constitués de deux périodes.

Comme vu précédemment dans les essais croisés, plusieurs mesures sont réalisées au cours du temps. Il est donc possible que la période influence la mesure de l'effet traitement. Ce phénomène est appelé l'effet période. Comme lors de la réalisation des modèles à effets fixes (*5*), il est possible de tenir compte de l'effet période. Cependant, dans ce cas-ci, la corrélation n'est pas incluse dans le modèle via un effet patient mais via la matrice de variance-covariance des résidus.

Le modèle est alors :

$$y_{ijk} = \mu + \alpha_{1,j} treat_j + \alpha_{2,k} p_k + \epsilon_{ijk}$$

$$i = 1, \dots, n \quad j = A, B \quad k = 1,2$$

avec

$\alpha_{2,k}$: effet période

L'effet période est considéré comme une variable qualitative, comme c'est généralement le cas dans les essais croisés AB/BA. p_k prendra la valeur de 0 si la mesure est réalisée à la période de référence et 1 sinon. Le paramètre $\alpha_{1,j}$ correspond à l'effet traitement.

Puisque dans les essais croisés les mesures sont répétées plusieurs fois chez un même patient, les observations ne sont pas indépendantes. La corrélation entre les observations peut être prise en compte en utilisant un modèle appelé, *covariance pattern model*. Dans ce type de modèle, la structure de corrélation n'est pas prise en compte par des effets aléatoires mais directement par la matrice **R**, matrice de variance des résidus.

$$\epsilon_{ijk} \sim N(0, \boldsymbol{R})$$

La matrice **R** pour un certain patient *i*, notée R_i, sera de dimension $K \times K$, où K correspond au nombre de périodes pendant lesquelles les mesures ont été récoltées pour ce patient. Via cette matrice **R**, plusieurs structures de dépendance peuvent être envisagées. Comme par exemple, des corrélations entre les données qui diminuent au fur et à mesure que le temps entre les mesures augmente (autorégressive), des structures de corrélation qui sont fonction de l'écart temporel (Toeplitz) ou encore aucune structure particulière (non structurée). Ci-dessous trois schémas illustrant une matrice R_i, dans le cas de 3 périodes, non structurée (1) ainsi que les deux types de structures souvent utilisés dans les modèles à Covariance Pattern, la structure Toeplitz (2) et la structure autorégressive (3) sont présentés :

$$(1) = \begin{pmatrix} {\sigma_1}^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & {\sigma_2}^2 & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & {\sigma_3}^2 \end{pmatrix} (2) = \begin{pmatrix} \sigma^2 & \sigma_1 & \sigma_2 \\ \sigma_1 & \sigma^2 & \sigma_1 \\ \sigma_2 & \sigma_1 & \sigma^2 \end{pmatrix} (3) = \begin{pmatrix} \sigma^2 & \sigma^2\rho & \sigma^2\rho^2 \\ \sigma^2\rho & \sigma^2 & \sigma^2\rho \\ \sigma^2\rho^2 & \sigma^2\rho & \sigma^2 \end{pmatrix}$$

Formule 1 : Structures de la matrice Ri, la matrice R pour un certain patient i, dans les modèles à Covariance Pattern. Les matrices ci-dessus représentent une matrice non structurée (1), une structure Toeplitz (2) et une structure autorégressive (3). Cet exemple illustre un cas où le nombre de périodes est égal à 3 (k=3).

Dans le cas d'une matrice non structurée pour une étude contenant 3 périodes (1), la matrice de variance-covariance est constituée d'un total de 6 paramètres à estimer (3 paramètres de variance ($\sigma_1^2,\ \sigma_2^2,\ \sigma_3^2$) et 3 paramètres de covariance ($\sigma_{12}, \sigma_{13}, \sigma_{23}$). Pour une matrice de structure Toeplitz (2), 3 paramètres sont à estimer (1 paramètre de variance (σ^2) et 2 paramètres de covariance (σ_1, σ_2)). Enfin, dans le cas d'une structure autorégressive (3), seulement 2 paramètres sont à estimer (1 paramètre de variance (σ^2) et le coefficient de corrélation, ρ, qui correspond à la corrélation entre deux observations séparées par une seule période).

La matrice **R** sera définie comme une matrice en blocs de dimension $R_i \times R_i$. Cela signifie qu'une matrice carrée, comme celles représentées ci-dessus, sera définie en fonction de la variable bloc choisie (les patients dans ce cas-ci). Les matrices R_i pour chaque patient seront situées sur la diagonale. Tandis que des 0 seront placés partout ailleurs puisque les observations provenant de deux patients différents sont considérées comme indépendantes.

$$\mathrm{R} = \begin{pmatrix} R_1 & 0 & \cdots \\ 0 & R_2 & \cdots \\ \cdots & \cdots & \cdots \end{pmatrix}$$

Formule 2 : Représentation de la matrice R dans le cas des modèles à covariance pattern des essais croisés

La forme matricielle de ce modèle $Y = X\alpha + \epsilon$ peut être représentée. En repartant des deux premiers patients de l'exemple de la *figure 7* avec $X_{i,1:2}$ la matrice design des effets fixes des deux premiers patients et en considérant la première période comme la période de référence et le salbutamol comme le traitement de référence alors la forme matricielle est :

$$X_{i,1:2} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix} \quad \alpha = (\mu, \alpha_{1,B}, \alpha_{2,2})'$$

A ce modèle, une interaction traitement-période pourrait être rajoutée. Cette interaction permet de prendre en compte le fait que l'effet du traitement peut être différent en fonction de la période à laquelle il est administré. Le modèle devient alors :

$$y_{ijk} = \mu + \alpha_{1,j} treat_j + \alpha_{2,k} p_k + \alpha_{3,jk} treat_j\, p_k + \epsilon_{ijk}$$

$$i = 1, \dots, n \quad j = A, B \quad k = 1,2$$

avec

$\alpha_{3,jk}$: effet d'interaction traitement-période

Ce deuxième modèle peut aussi être réécrit sous forme matricielle $Y = X\alpha + \epsilon$, pour les deux premiers patients de l'étude du formotérol et du salbutamol :

$$X_{i,1:2} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix} \quad \alpha = (\mu, \alpha_{1,B}, \alpha_{2,2}, \alpha_{3,B2})'$$

3.2.3.3 Effet carry over

Comme décrit précédemment, une des complications possibles dans les analyses des essais croisés est la présence d'un potentiel effet *carry over*. Pour rappel, un effet *carry over* est un phénomène qui peut apparaître lorsque les effets d'un traitement sont prolongés dans le temps. Les effets mesurés ne sont alors plus les effets individuels mais bien les effets cumulés des traitements. Afin d'estimer l'effet *carry over*, il est nécessaire d'inclure dans le modèle la notion de *séquence,* décrite au point 3.2.1. Pour rappel, la séquence correspond à l'ordre dans lequel le patient a reçu les traitements et est notée *s*. Le modèle est construit à partir d'un modèle simple, contenant l'effet traitement et l'effet période, auquel sont ajoutés un effet *carry over* et un effet aléatoire patient.

L'observation y_{isjk} du patient *i* appartenant à la séquence *s* ayant reçu la traitement *j* à la période *k* est modélisée comme :

$$y_{isjk} = \mu + \alpha_{1,j} treat_j + \alpha_{2,k} p_k + \alpha_{3,s} c_s + \beta_{1,i} + \epsilon_{isjk}$$

$$i = 1, \dots, n \;\; \text{s} = 1,2 \;\; j = A, B \;\; k = 1,2$$

avec

$\alpha_{3,s}$: effet de la séquence *s* ou effet *carry over*

$\beta_{1,i}$: effet du patient *i*

$c_{3,s}$ est une variable indicatrice qui correspond à la séquence à laquelle appartient le patient. L'indicateur $c_{3,s}$ prendra la valeur de 0 pour la séquence de référence, par exemple A/B, et 1 pour la séquence B/A (*voir figure 7*). L'estimateur $\hat{\alpha}_{3,s}$ estime donc la différence entre les deux séquences. S'il existe une différence entre les séquences A/B et B/A alors l'ordre des traitements a une importance et cela peut être un signe de la présence d'un effet *carry over*. Ce modèle permet donc d'estimer l'effet *carry over* tout en tenant compte grâce au modèle de l'effet période et de l'effet patient. La période est considérée comme une variable qualitative. p_k prendra la valeur de 0 si la mesure est effectuée à la période de référence et 1 sinon. L'effet traitement est considéré comme précédemment. $treat_j$ prendra la valeur de 0 si la mesure est réalisée après l'administration du traitement de référence et 1 sinon.

En partant de la *figure 7* pour les deux premiers patients et en considérant que la séquence salbutamol/formoterol est la séquence de référence, la forme matricielle $Y = X\alpha + Z\beta + \epsilon$ de ce modèle s'écrit :

$$X_{i,1:2} = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \end{pmatrix} \quad \alpha = \left(\mu, \alpha_{1,B}, \alpha_{2,2}, \alpha_{3,B/A}\right)' \textbf{ et } Z_{i,1:2} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \quad \beta = (\beta_{1,1}, \beta_{1,2})'$$

Comme précédemment, $\hat{\beta}_{1,i}$ correspond donc à la déviation de l'intercept du patient *i* par rapport à la moyenne de tous les patients. Les $\beta_{1,i}$ sont considérés comme indépendants et identiquement distribués, de distribution normale :

$$\beta_{1,i} \sim N(0, {\sigma_x}^2)$$

Le paramètre ${\sigma_x}^2$ est la variance entre les différents patients. Senn propose de construire un modèle dans le but est de séparer la variance en deux sources : la variance inter-séquence et la variance inter-patient appartenant à la séquence *s* (Senn S. , 2019). Cependant, ce modèle n'a pas pu être reproduit à partir des données de l'étude du salbutamol et du formoterol.

La distribution des termes d'erreur reste :

$$\epsilon_{isjk} \sim N(0, \sigma^2)$$

Afin de mieux comprendre l'estimation de cet effet *carry over*, ce modèle est appliqué sur les données de l'étude des traitements formoterol et salbutamol. L'estimation de l'effet *carry*

over est de -7.2 l/min avec une statistique du test de Wald de 0.18. Ce test a comme hypothèse nulle que l'effet *carry over* est égal à 0 donc qu'il n'y pas d'effet *carry over* dans cette étude et comme hypothèse alternative que cet effet est différent de 0 :

$$H0: \alpha_{3,s} = 0$$

$$H1: \alpha_{3,s} \neq 0$$

Le résultat du test étant non significatif, l'hypothèse nulle n'est pas rejetée et la conclusion est qu'aucun effet *carry over* n'a été mis en évidence. Ces résultats sont donc identiques à ceux obtenus lors du test t réalisé au point 3.1.4.

Bien qu'aucun effet *carry over* n'ait été mis en évidence jusqu'ici, il ne faut pas oublier le fait que cet effet soit particulièrement difficile à estimer. Comme vu précédemment, il est extrêmement difficile de mesurer uniquement l'effet *carry over*, sans que cette estimation ne soit biaisée par d'autres facteurs. En effet, dans le modèle ci-dessus, la présence ou non d'un effet *carry over* a été testée sur base de la différence entre les séquences. Il est clair que cette différence peut être impactée par de nombreux facteurs comme l'effet de la période, du patient, d'une interaction patient-période ou encore d'une interaction traitement-patient. Bien que certains de ces facteurs aient été pris en compte par le modèle, il est possible que la différence entre les séquences ne reflète pas uniquement la présence d'un effet *carry over*. Ainsi l'estimation réalisée de l'effet *carry over* reste approximative. Dans ce travail seul un modèle simple est présenté, il faut cependant garder à l'esprit que de nombreux auteurs discutent, encore à l'heure actuelle, de la mesure et l'analyse de cet effet *carry over* dans le but d'améliorer son estimation.

3.3 Analyses statistiques des essais *stepped wedge*

Les essais *stepped wedge* sont considérés comme des essais croisés unidirectionnels (Twisk, 2021). En effet dans ce type d'essais, les groupes d'individus commencent tous par une période de contrôle. Un traitement est ensuite implémenté progressivement dans chacun des groupes au cours des différentes périodes.

La figure suivante illustre la structure générale des essais *stepped wedge* pour un essai constitué de 8 groupes et 5 périodes (Li, Hughes, Hemming, & Taljaard, 2021):

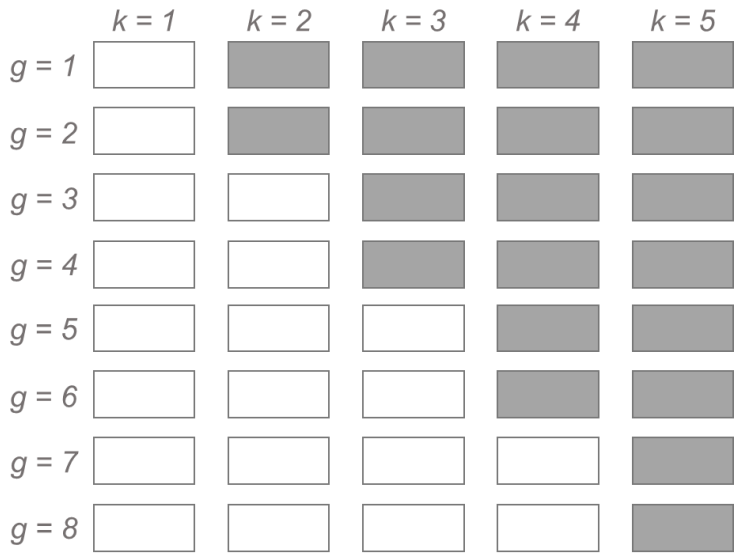

Figure 8 : Structure d'un essai stepped wedge avec k représentant les périodes (k=1,…,5) et g représentant les groupes (g=1,…,8).

Dans cet exemple de structure des essais *stepped wedge*, 8 groupes différents, notés $g = 1, \ldots, 8$, sont soumis à la mise en place d'un traitement. Ce traitement est progressivement installé au cours de 5 périodes, notées $k = 1 \ldots, 5$. Le traitement est représenté par un rectangle gris tandis que le contrôle par un rectangle blanc. La première période $(k = 1)$ constitue une période de référence au cours de laquelle le traitement n'a pas encore été implémenté dans les groupes.

Dans ce type d'essais, il est important de prendre en considération le temps (Twisk, 2021). Dans les essais *stepped wedge*, la durée des études est généralement importante. Comme expliqué lors de l'analyse des essais croisés, le temps peut avoir un effet sur les individus des différents groupes. De plus, la durée des périodes varie en fonction des études. Dans la littérature, certaines études ont des périodes de quelques semaines tandis que d'autres de plusieurs mois. Généralement, la durée des périodes dépend du temps nécessaire à la préparation et l'organisation de la mise en place du traitement dans les différents groupes.

Le design des essais *stepped wedge* fait apparaitre une notion importante, appelée « séquence », qui n'a pas la même signification que celle vue lors de l'analyse des essais croisés. Une séquence est définie comme le nombre de passage de l'état de contrôle à la

situation de traitement. Dans la *figure 8*, il existe donc 4 séquences distinctes. Les différents groupes vont chacun leur tour passer de l'état de contrôle à l'état de traitement.

Deux idées principales se cachent derrière l'analyse des essais *stepped wedge* (Twisk, 2021). La première est, comme dans les essais croisés, la comparaison des observations récoltées chez un même individu ou dans un même groupe d'individus. La deuxième est la comparaison entre la situation de contrôle et la situation de traitement permise grâce à la configuration des essais *stepped wedge*. En effet, dans ce type d'essais, le traitement peut être comparé à chacune des étapes au contrôle à partir de la deuxième période (voir *figure 8*). La difficulté ici est de réussir à mesurer ces deux aspects des essais *stepped wedge* dans une même analyse.

Plusieurs types d'essais *stepped wedge* existent. Ces derniers sont *cross-sectional*, *closed cohort* et *open cohort*. Les essais *stepped wedge cross-sectional* sont des essais pour lesquels différents individus sont observés dans chaque groupe à chaque période. Une seule observation par individu est donc récoltée soit dans la situation de contrôle soit dans la situation de traitement. Les essais *closed cohort* sont des essais pour lesquels les observations par individu sont répétées avec des mesures prises dans les différentes périodes donc sous contrôle et sous traitement. Les essais *stepped wedge open cohort* sont des essais dans lesquels des individus peuvent s'ajouter à chaque période au cours de l'étude.

L'utilisation d'un modèle mixte pour analyser les essais *stepped wedge* offre plusieurs avantages (Twisk, 2021). Certains avantages sont communs à ceux vus lors de l'analyse des essais croisés. Un premier avantage est l'ajustement des observations répétées et corrélées chez un même patient (*closed cohort* et *open cohort*). Un deuxième avantage est l'ajustement des observations corrélées entre les sujets venant d'un même groupe. En effet dans les essais *stepped wedge,* le traitement est mis en place progressivement dans plusieurs centres de soins différents (ou hôpitaux, écoles, …) amenant donc plusieurs groupes d'individus. Les modèles mixtes permettent donc de prendre en considération plusieurs effets qui ont potentiellement un impact sur l'effet du traitement en une seule et même analyse.

Il est important de faire la différence entre les trois types d'essais puisque les analyses statistiques sont spécifiques à chacun des types d'essais *stepped wedge*. Dans ce travail, les *stepped wedge* de types *cross-sectional* et *closed cohort* seront présentés*.* Les essais *stepped wedge open cohort* ne seront pas étudiés.

Les modèles présentés ci-dessous proviennent principalement de l'article « Mixed-effects models for the design and analysis of stepped wedge cluster randomized trials: An overview » datant de 2021 (Li, Hughes, Hemming, & Taljaard, 2021).

3.3.1 Cross-sectional

Les essais *stepped wedge cross-sectional* sont des essais dans lesquels la mesure pour un individu *i* dans un groupe *g* n'est réalisée qu'une seule fois à une certaine période *k*, de contrôle ou de traitement (Patterson, Leland, Mormer, & Palmerd, 2022). Autrement dit à chacune des périodes, de nouveaux individus sont observés pour chaque groupe. Des

exemples de situations dans lesquelles ces essais sont utilisés pourraient être l'évaluation d'une nouvelle technique chirurgicale comparée à une ancienne, d'une nouvelle technique de soins post-accouchement comparée au standard ou encore la mise en place d'un vaccin. Dans ce cas, la chirurgie, l'accouchement ou le vaccin ne sont réalisés qu'une seule fois sur les individus. Les données seront alors récoltées par la suite au cours d'une séance de suivi médical. Une étude réalisée en 2013 avec pour objectif d'améliorer l'accompagnement et la communication entre médecins et patients dans le cadre d'un accouchement a été réalisée à l'aide du design *stepped wedge cross-sectional* (Bashour, et al., 2013). Des formations dédiées au personnel de l'hôpital étaient donc organisées et implémentées progressivement dans les différents hôpitaux. Les femmes enceintes recevaient soit le suivi standard soit les nouvelles méthodes de communication en fonction du moment de leur accouchement et du moment de l'implémentation de la nouvelle méthode dans l'hôpital en question. Ces dernières étaient alors questionnées dans les deux semaines après leur accouchement et les données étaient récoltées à ce moment-là. Il serait également envisageable d'imaginer une situation dans laquelle il n'est pas possible de récolter des données pour tous les patients tout au long de l'étude. Dans ce cas, les patients seraient alors suivis pendant la même période mais pour chacun des patients, une période serait définie aléatoirement pour récolter les données. Cependant, les études utilisant des essais *stepped wedge cross-sectional* dans le cas où les données de chaque individu ne peuvent pas être récoltées à chacune des périodes sont moins fréquemment rencontrées dans la littérature. Généralement dans les essais *stepped wedge cross-sectional,* le contrôle ou le traitement est appliqué à une certaine période sur un individu et évalué une seule fois après l'administration. Les individus ne sont donc pas suivis au cours du temps.

A partir de la *figure 8*, une représentation des trois premiers individus du premier groupe ($g = 1$) et du deuxième groupe ($g = 2$), dans le cas d'un essai *steped wedge cross-sectional* peut être imaginée comme suit :

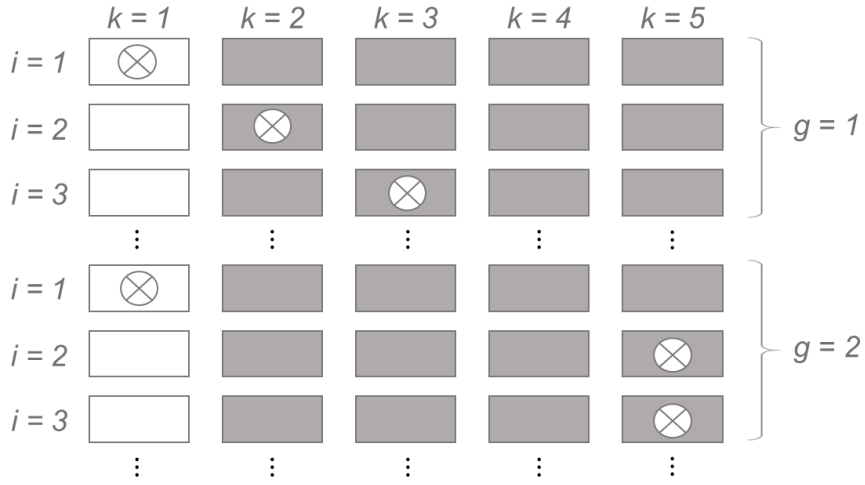

Figure 9 : Représentation des 3 premiers individus du premier groupe et du deuxième groupe dans le cas d'un essai stepped wedge cross-sectional (exemple fictif). L'implémentation du traitement dans le groupe est représentée par un rectangle gris et la situation de contrôle par un rectangle blanc. Les cases marquées par une cible représentent la période à laquelle la mesure a été réalisée.

Pour rappel, dans le cas d'un essai *stepped wedge cross-sectional*, une seule mesure par individu est récoltée. A la différence des essais croisés, il n'est pas nécessaire de considérer la variabilité des mesures intra-patient étant donné qu'une seule observation est disponible par patient. Il est considéré ici que les mesures sont réalisées directement après l'administration du contrôle ou du traitement au patient.

3.3.1.1 Modèle simple

Ce premier modèle est un modèle constitué de deux effets, un effet fixe traitement et un effet aléatoire groupe. Étant donné que plusieurs individus proviennent du même groupe, il est possible que les observations provenant d'un même groupe soient corrélées. Cette corrélation est donc prise en compte par l'effet aléatoire groupe. Le modèle pour l'observation du patient *i* appartenant au groupe *g* pour la situation de traitement *j* est :

$$y_{igj} = \mu + \alpha_{1,j} treat_j + \beta_{1,g} + \epsilon_{igj} \qquad (7)$$

$$i = 1, \dots, n \quad g = 1, .., G \quad j = 0,1$$

avec

μ : terme d'intercept

$\alpha_{1,j}$: effet traitement

$\beta_{1,g}$: effet groupe

Dans ce modèle, $treat_j$ est une variable indicatrice qui correspond à la situation du patient, situation de contrôle ou de traitement. Cette variable $treat_j$ prendra la valeur de 1 si le groupe *g* au temps *k* est dans la situation de traitement $(j = 1)$ et 0 si le groupe est dans la situation de contrôle $(j = 0)$. Le paramètre $\alpha_{1,j}$ représente la différence du traitement par rapport à la situation de contrôle. Dans ce cas, $\mu + \alpha_{1,j} treat_j$ correspond à la valeur moyenne de Y pour les individus dans la situation de traitement. L'intercept, μ représente la réponse moyenne de base, avant l'implémentation du traitement donc pour la situation de contrôle.

L'effet groupe, $\beta_{1,g}$, est considéré ici comme un effet aléatoire. Il a donc une distribution propre, supposée normale. Comme vu dans les analyses des essais croisés, un $\hat{\beta}_{1,g}$ peut être prédit pour chacun des groupes et correspond à la déviation de l'intercept du groupe *g* par rapport à la moyenne de tous les groupes. Pour rappel, ϵ_{igj} représente le terme d'erreur lié à la régression pour chacune des observations de chacun des patients de chacun des groupes. Le terme ϵ_{igj} correspond donc à l'erreur du patient *i* appartenant au groupe *g* pour la situation de traitement *j.* Ils sont considérés comme indépendants, identiquement distribués et ayant une distribution normale :

$$\beta_{1,g} \sim N(0, {\sigma_g}^2)$$

$$\epsilon_{igj} \sim N(0, \sigma^2)$$

Les termes $\beta_{1,g}$ et ϵ_{igj} sont supposés indépendants. En ce qui concerne l'effet aléatoire groupe, la composante de variance correspondante, notée ${\sigma_g}^2$, représente la variation qui existe entre les groupes.

Ce premier modèle peut être réécrit sous une forme matricielle $Y = X\alpha + Z\beta + \epsilon$. A partir de la *figure 9*, en considérant $X_{g,1:2}$ et $Z_{g,1:2}$, les matrices de designs des effets fixes et aléatoires pour les premier et deuxième groupes, respectivement, la forme matricielle des trois patients du premier et du deuxième groupes est telle que :

$$X_{g,1:2} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & 0 \\ 1 & 1 \\ 1 & 1 \\ \vdots & \vdots \end{pmatrix} \alpha = (\mu, \alpha_{1,1})' \textbf{ et } Z_{g,1:2} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ \vdots & \vdots \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ \vdots & \vdots \end{pmatrix} \beta = (\beta_{1,1}, \beta_{1,2})'$$

Comme vu précédemment dans les essais croisés, la variance des observations individuelles dans ce modèle à effet aléatoire se calcule comme la somme de toutes les composantes de variance et est égale à :

$$Var(y_{igj}) = {\sigma_g}^2 + \sigma^2$$

Dans ce modèle, les observations provenant de deux groupes différents sont considérées comme indépendantes. Cependant, les observations d'un même groupe (g = g') ne sont pas considérées comme indépendantes avec :

$$Cov(y_g, y_{g'}) = {\sigma_g}^2 \qquad \text{pour } g = g'$$

La corrélation entre deux individus, *i* et *j*, venant d'un même groupe (g = g') sera obtenue comme suit (Li, Hughes, Hemming, & Taljaard, 2021):

$$Corr(y_{ig}, y_{jg'}) = \rho = \frac{{\sigma_g}^2}{({\sigma_g}^2 + \sigma^2)} \qquad \text{pour } g = g'$$

3.3.1.2 Modèle de Hussey et Hughes

Dans le premier modèle proposé (*7*), le fait que les données soient récoltées au cours du temps n'est pas pris en compte. Dans les essais croisés, le temps était considéré comme ayant une influence possible sur l'individu (*modèle (5*)). Dans un essai *stepped wedge,* cet impacte pourrait être d'autant plus important étant donné que généralement les essais *stepped wedge* sont constitués de plusieurs périodes, comparés aux essais croisés AB/BA. Il est donc nécessaire de rajouter un paramètre dans le modèle de base afin de considérer un potentiel impact des différentes périodes sur les groupes d'individus. Ce modèle est le modèle qui a été proposé par Hussey et Hughes (Li, Hughes, Hemming, & Taljaard, 2021).

L'observation y_{igkj} correspondant au patient *i* venant du groupe *g* à la période *k* pour la situation de traitement *j* est modélisée comme :

$$y_{igkj} = \mu + \alpha_{1,j} treat_j + \alpha_{2,k} p_k + \beta_{1,g} + \epsilon_{igkj} \quad (8)$$
$$i = 1, \dots, n \quad g = 1, .., G \quad k = 1, \dots, K \quad j = 0,1$$

avec

$\alpha_{2,k}$: effet période

et la distribution des termes $\beta_{1,g}$ et ϵ_{igkj} est supposée normale :

$$\beta_{1,g} \sim N(0, {\sigma_g}^2)$$
$$\epsilon_{igkj} \sim N(0, \sigma^2)$$

La composante de variance ${\sigma_g}^2$ correspond à la variance entre les groupes. σ^2 représente la variance à l'intérieur d'un groupe.

Dans les analyses des essais croisés, le temps a été décrit comme une variable considérée soit comme qualitative soit comme quantitative (voir point 3.2.2.3).

Dans le cas où la variable période est considérée comme qualitative alors l'effet période est en réalité un vecteur ${\alpha_{2,k}}'$. A partir de la *figure 8*, ce dernier peut être décomposé comme suit en considérant la première période comme la période de référence :

$${\alpha_{2,k}}' p_k = \alpha_{2,k=2}\, p_2 + \alpha_{2,k=3}\, p_3 + \alpha_{2,k=4}\, p_4 + \alpha_{2,k=5}\, p_5$$

L'effet fixe période est donc constitué de constantes et de dummy variables. Les paramètres, $\alpha_{2,k}$ avec $k = 2, \dots, 5$, correspondent à la différence entre la mesure à la période *k* et la mesure initiale réalisée à la période de référence. Les dummy variables, notées p_k avec $k = 2, \dots, 5$, prennent comme valeurs 0 ou 1 en fonction de la période à laquelle l'observation a été réalisée (Twisk, 2021). La période n°1 a été considérée comme la période de référence. L'estimation de l'effet de la période de référence sera reprise dans le terme d'intercept. Quatre paramètres $\alpha_{2,k}$, avec $k = 2, \dots, 5$, seront donc estimés.

Dans le cas où cette variable est considérée comme quantitative, alors le paramètre α_k correspondra à l'effet période et p_k représentera la période. Dans ce cas à partir de l'illustration de la *figure 8*, p_k prendre des valeurs de 1 à 5 en fonction de la période à laquelle l'observation a été récoltée. Un seul paramètre sera alors estimé. En considérant la variable période comme une variable quantitative, une relation linéaire est supposée entre la variable réponse et la variable période.

Au vu du design des essais *stepped wedge* et étant donné que les périodes sont parfois de longue durée, les intervalles de temps entre les observations récoltées peuvent être forts différents. Dans ce cas, il est préférable d'utiliser des dummy variables et d'estimer un paramètre $\alpha_{2,k}$ pour chacune des périodes de l'étude avec une période considérée comme la référence. Dans ce travail, cette variable sera principalement considérée comme qualitative.

En reprenant les trois premiers patients du premier groupe et du deuxième groupe de la *figure 9* et en considérant $X_{g,1:2}$ et $Z_{g,1:2}$, les matrices designs des effets fixes et aléatoires pour le premier groupe et le deuxième groupe, respectivement, la forme matricielle $Y = X\alpha + Z\beta + \epsilon$ de ce modèle est :

$$X_{g,1:2} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix} \quad \alpha = \begin{pmatrix} \mu \\ \alpha_{1,1} \\ \alpha_{2,2} \\ \alpha_{2,3} \\ \alpha_{2,4} \\ \alpha_{2,5} \end{pmatrix} \quad \textbf{et} \quad Z_{g,1:2} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ \vdots & \vdots \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ \vdots & \vdots \end{pmatrix} \quad \beta = (\beta_{1,1}, \beta_{1,2})'$$

A partir de ce modèle, Hussey et Hughes définissent un modèle des moyennes des groupes. Dans ce nouveau modèle, ce n'est plus la mesure récoltée pour chaque patient qui est modélisée mais la moyenne des observations de tous les patients de chacun des groupes.

La modélisation de la <u>moyenne</u> y_{gkj} du groupe *g* à la période *k* appartenant à la situation *j* est :

$$y_{gkj} = \mu + \alpha_{1,j} treat_j + \alpha_{2,k}{}' p_k + \beta_{1,g} + \epsilon_{gkj} \qquad (9)$$

$$g = 1, .., G \quad k = 1, \dots, K \quad j = 0,1$$

Dans ce deuxième modèle, l'erreur ϵ_{gkj} est alors considérée, avec n_g correspondant au nombre d'individus échantillonnés par groupe par période, comme :

$$\epsilon_{gkj} = \frac{\sum_{i=1}^{n_g} \epsilon_{igkj}}{n_g}$$

Les auteurs considèrent ici le nombre d'observations récoltées par groupe par période, n_g, comme étant constant.

La distribution de ϵ_{gkj} est alors :

$$\epsilon_{gkj} \sim N(0, {\sigma_{wg}}^2)$$

Dans ce contexte, ${\sigma_{wg}}^2$ correspond à σ^2/n_g. Les termes ϵ_{igkj} et ϵ_{gkj} sont considérés comme étant indépendants de $\beta_{1,g}$.

Pour rappel, la variance de y_{igkj} correspond à :

$$Var(y_{igkj}) = {\sigma_g}^2 + \sigma^2$$

Tandis que la variance de y_{gkj} est égale à :

$$Var(y_{gkj}) = {\sigma_g}^2 + {\sigma_{wg}}^2$$

Ces notions seront utilisées plus tard dans le travail lors de la présentation des calculs de la taille d'échantillon des essais *stepped wedge* (voir point 4.2.1).

3.3.1.3 *Effet d'interaction période-groupe*

Dans l'analyse des essais croisés, l'effet période était considéré comme différent en fonction des patients. Dans les essais *stepped wedge*, il serait donc raisonnable de penser que l'effet période n'est pas le même pour tous les groupes. Un effet aléatoire représentant l'interaction entre l'effet groupe et l'effet période peut donc être ajouté au modèle précédent (*8*). Le modèle devient alors :

$$y_{igkj} = \mu + \alpha_{1,j} treat_j + \alpha_{2,k}' p_k + \beta_{1,g} + \beta_{2,gk} + \epsilon_{igkj} \qquad (10)$$
$$i = 1, \dots, n \quad g = 1, .., G \quad k = 1, \dots, K \quad j = 0,1$$

avec

$\beta_{2,gk}$: effet d'interaction groupe-période

$\hat{\beta}_{1,g}$ correspond à la déviation de l'intercept du groupe *g* par rapport à la moyenne de tous les groupes. Étant donné que la variable période est considérée comme qualitative, un $\hat{\beta}_{2,gk}$ sera prédit pour chacune des périodes de chacun des groupes. Les $\beta_{2,gk}$ sont indépendants, identiquement distribués et ayant une distribution normale :

$$\beta_{2,gk} \sim N(0, {\sigma_{gk}}^2)$$

La forme matricielle $Y = X\alpha + Z\beta + \epsilon$ de ce modèle est la suivante :

$$X_{g,1:2} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix} \quad \alpha = \begin{pmatrix} \mu \\ \alpha_{1,1} \\ \alpha_{2,2} \\ \alpha_{2,3} \\ \alpha_{2,4} \\ \alpha_{2,5} \end{pmatrix}$$

et

$$Z_{g,1:2} = \begin{pmatrix} 1 & 1 & 0 & 0 \cdots 0 & 0 \cdots 0 \\ 1 & 0 & 1 & 0 \cdots 0 & 0 \cdots 0 \\ 1 & 0 & 0 & 1 \cdots 0 & 0 \cdots 0 \\ \vdots & \vdots & \vdots & \vdots \quad \vdots & \vdots \quad \vdots \\ 0 & 0 & 0 & 0 \cdots 1 & 1 \cdots 0 \\ 0 & 0 & 0 & 0 \cdots 1 & 0 \cdots 1 \\ 0 & 0 & 0 & 0 \cdots 1 & 0 \cdots 1 \\ \vdots & \vdots & \vdots & \vdots \quad \vdots & \vdots \quad \vdots \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_{1,1} \\ \beta_{2,11} \\ \beta_{2,12} \\ \beta_{2,13} \\ \cdots \\ \beta_{1,2} \\ \beta_{2,21} \\ \cdots \\ \beta_{2,25} \end{pmatrix}$$

Les auteurs considèrent les deux effets aléatoires, $\beta_{1,g}$ et $\beta_{2,gk}$,comme étant indépendants (Li, Hughes, Hemming, & Taljaard, 2021). Cependant, il serait possible de ne pas les considérer comme étant indépendants. Cette possibilité est détaillée par la suite.

Dans les modèles tenant compte de l'effet période, la corrélation entre deux observations d'un même groupe n'est plus la même, comparée à celle détaillée dans le modèle simple (voir point 3.3.1.1). En effet, dans ce nouveau modèle, deux types de corrélation sont à considérer. La première est la corrélation '*within-period*', notée ρ_{wp}, qui correspond à la corrélation entre deux observations d'un même groupe *g* collectées au cours de la même période *k* ($k = k'$). La seconde est la corréaltion *'between-period'*, notée ρ_{bp}, qui correspond à la corrélation entre deux observations d'un même groupe *g* collectées à deux périodes différentes ($k \neq t$).

$$\rho_{wp} = \frac{{\sigma_g}^2 + {\sigma_{gk}}^2}{({\sigma_g}^2 + {\sigma_{gk}}^2 + \sigma^2)}$$

$$\rho_{bp} = \frac{{\sigma_g}^2}{({\sigma_g}^2 + {\sigma_{gk}}^2 + \sigma^2)}$$

A partir du modèle (10), les auteurs envisagent un modèle dans lequel la variable période est considérée comme une variable quantitative. Ce modèle impose une relation linéaire entre la variable réponse et le temps. Ce modèle est appelé un modèle à coefficients aléatoires car le terme aléatoire est associé à une variable continue. Sur base de (10), le modèle à coéfficients aléatoires s'écrit comme suit :

$$y_{igkj} = \mu + \alpha_{1,j} treat_j + \alpha_{2,k} p_k + \beta_{1,g} + \beta_{2,gk} p_k + \epsilon_{igkj}$$

La forme matrcielle de ce modèle est :

$$X_{g,1:2} = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 2 \\ 1 & 1 & 3 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 1 \\ 1 & 1 & 5 \\ 1 & 1 & 5 \\ \vdots & \vdots & \vdots \end{pmatrix} \alpha = (\mu, \alpha_{1,1}, \alpha_{2,k})' \textbf{ et } \quad Z_{g,1:2} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 2 & 0 & 0 \\ 1 & 3 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 5 \\ 0 & 0 & 1 & 5 \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix} \beta = (\beta_{1,1}, \beta_{2,1k}, \beta_{1,2}, \beta_{2,2k})'$$

$\hat{\beta}_{2,gk}$ correspond à l'effet période spécifique à chaque groupe. Dans ce modèle, les coéfficients aléatoires $\beta_{1,g}$ et $\beta_{2,gk}$ ne sont pas considérés comme étant indépendants. La distribution de ces coefficients est alors une normale bivariée :

$$\begin{pmatrix} \beta_{1,g} \\ \beta_{2,gk} \end{pmatrix} \sim N(0, \boldsymbol{G})$$

avec

$$\boldsymbol{G} = \begin{pmatrix} {\sigma_g}^2 & \sigma_{g,gk} \\ \sigma_{gk,g} & {\sigma_{gk}}^2 \end{pmatrix}$$

Les composantes de variance ${\sigma_g}^2$ et ${\sigma_{gk}}^2$ représentent la variance de l'effet de groupe et la variance de l'effet d'interaction groupe-période, respectivement. Le terme $\sigma_{g,gk}$ représente la covariance entre les deux coefficients aléatoires. Dans ce modèle, une déviation par rapport

au terme d'intercept sera estimée pour chacun des groupes comme dans le modèle de base. De plus, une pente pour chacun des groupes sera également estimée permettant donc au temps d'avoir un effet différent en fonction du groupe. Cependant, ces modèles n'ont pas encore été réellement étudiés dans le cas des essais *stepped wedge* (Li, Hughes, Hemming, & Taljaard, 2021).

Les auteurs envisagent également un troisième modèle à effets aléatoires. Dans ce modèle, la variable période est considérée comme qualitative. Dans ce troisième modèle, la corrélation qui existe entre les mesures réalisées dans un même groupe est reprise par l'effet aléatoire groupe-période (Li, Hughes, Hemming, & Taljaard, 2021).

Le modèle s'écrit alors comme suit :

$$y_{igkj} = \mu + \alpha_{1,j} treat_j + \alpha_{2,k}' p_k + \beta_{1,gk} + \epsilon_{igkj} \qquad (11)$$

La corrélation entre les données est prise en considération grâce à l'effet aléatoire d'interaction groupe-période, la distribution des $\beta_{1,gk}$ est :

$$\beta_{1,gk} \sim N(0, \boldsymbol{G})$$

La matrice **G** se présente comme la matrice **R** vue précédemment dans les essais croisés (point 3.2.3.2). En reprenant l'illustration de la *figure 8*, la matrice **G** pour un seul groupe, notée $\boldsymbol{G}_{g,1}$, considérée comme non structurée est telle que :

$$G_{g,1} = \begin{pmatrix} {\sigma_{g1}}^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} & \sigma_{15} \\ \sigma_{21} & {\sigma_{g2}}^2 & \sigma_{23} & \sigma_{24} & \sigma_{25} \\ \sigma_{31} & \sigma_{32} & {\sigma_{g3}}^2 & \sigma_{34} & \sigma_{35} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & {\sigma_{g4}}^2 & \sigma_{45} \\ \sigma_{51} & \sigma_{52} & \sigma_{53} & \sigma_{54} & {\sigma_{g5}}^2 \end{pmatrix}$$

Formule 3 : Illustration d'une matrice G non structurée dans les essais stepped wedge cross-sectional.

Si la matrice est laissée non structurée alors le nombre de paramètres à estimer peut être élevé. En effet pour cette matrice réalisée sur base de la *figure 8*, 15 paramètres sont à estimer (5 paramètres de variance (σ_{gk}^2, $k = 1, \dots, 5$) et 10 paramètres de covariances ($\sigma_{12}, \dots, \sigma_{45}$), notés σ_{kt}).

La matrice **G** est constituée d'une structure en blocs, comme celle présentée ci-dessus, pour chacun des groupes sur la diagonale. Des 0 sont placés partout ailleurs étant donné que les observations provenant de deux groupes différents sont considérées comme non corrélées. σ_{gk}^2, avec $k = 1, \dots, 5$, sont donc les composantes de variance de l'effet d'interaction groupe-période et la covariance entre deux périodes ($k \neq t$) est notée σ_{kt}.

La covariance entre deux observations de deux individus différents provenant du même groupe ($g = g'$) à la même période ($k = k'$) correspondra à :

$$Cov(y_{gk}, y_{g'k'}) = {\sigma_{gk}}^2 \qquad \text{pour } (k = k')$$

Tandis que la covariance entre deux observations d'individus différents provenant d'un même groupe ($g = g'$) mais à deux périodes différentes ($k \neq t$) correspondra à :

$$Cov(y_{gk}, y_{g't}) = \sigma_{kt} \qquad \text{pour } (k \neq t)$$

σ_{kt} représente donc la covariance qui existe entre deux observations d'un même groupe réalisées à deux périodes différentes avec $k \neq t$. Deux observations qui ne proviennent pas du même groupe sont considérées comme non corrélées.

Lorsqu'une structure de corrélation de type *Toeplitz* est envisagée, alors la matrice $\boldsymbol{G}_{g,1}$, pour un groupe, se présente comme :

$$G_{g,1} = \begin{pmatrix} {\sigma_{gk}}^2 & \sigma_1 & \sigma_2 & \sigma_3 & \sigma_4 \\ \sigma_1 & {\sigma_{gk}}^2 & \sigma_1 & \sigma_2 & \sigma_3 \\ \sigma_2 & \sigma_1 & {\sigma_{gk}}^2 & \sigma_1 & \sigma_2 \\ \sigma_3 & \sigma_2 & \sigma_1 & {\sigma_{gk}}^2 & \sigma_1 \\ \sigma_4 & \sigma_3 & \sigma_2 & \sigma_1 & {\sigma_{gk}}^2 \end{pmatrix}$$

Formule 4 : Illustration d'une structure Toeplitz d'une matrice G dans les essais stepped wedge cross-sectional.

Avec une structure *Toeplitz,* le nombre de paramètres à estimer est moins élevé comparé à une matrice non structurée. Dans ce cas-ci, le nombre de paramètres à estimer se réduit à 5 paramètres (1 paramètre da variance (${\sigma_{gk}}^2$) et 4 paramètres de covariance ($\sigma_1, \dots, \sigma_4$)).

Le nombre de paramètres à estimer dans le cas d'une matrice non structurée peut vite devenir important lorsque le nombre de périodes est élevé. L'utilisation d'une matrice de type *Toeplitz* voir d'une matrice autorégressive permettrait de diminuer le nombre de paramètres à estimer. Il est nécessaire de bien garder à l'esprit, comme déjà mentionné précédemment, qu'une structure *Toeplitz* ou une structure autorégressive implique une corrélation décroissante en fonction de la distance entre les périodes (Li, Hughes, Hemming, & Taljaard, 2021). Il est donc nécessaire que les observations d'une même période soient récoltées plus ou moins au même moment pour que les intervalles entre les périodes restent plus ou moins constants. Cependant, au vu du design des essais *stepped wedge*, il semblerait que dans la pratique récolter les données à intervalles constants soit parfois compliqué.

3.3.1.4 Effet d'interaction traitement-groupe

Une interaction intéressante à étudier est l'interaction traitement-groupe. Etant donné que le traitement est mis en place dans chacun des groupes (hôpitaux, centres de soins…), il est possible que l'effet du traitement soit spécifique à chacun des groupes.

Pour modéliser cette interaction, un effet aléatoire est ajouté au modèle de base (*8*) proposé par Hussey et Hughes. Le modèle devient alors :

$$y_{igkj} = \mu + \alpha_{1,j} treat_j + \alpha_{2,k}{}' p_k + \beta_{1,g} + \beta_{2,gj} treat_j + \epsilon_{igkj}$$

$$i = 1, \dots, n \quad g = 1, .., G \quad k = 1, \dots, K \quad j = 0,1$$

avec

$\beta_{2,gj}$: effet traitement spécifique au groupe *g*

et l'effet aléatoire traitement-groupe est supposé suivre une distribution normale :

$$\beta_{2,gj} \sim N(0, {\sigma_{gj}}^2)$$

La variable période est considérée comme qualitative. Un $\hat{\beta}_{2,gj}$ peut être prédit pour chacun des groupes. Les $\beta_{2,gj}$ sont considérés comme indépendants, identiquement distribués et ayant une distribution normale. Les effets aléatoires, $\beta_{1,g}$ et $\beta_{2,gj}$, ne sont pas considérés comme indépendants. Dans ce cas, ce modèle suppose :

$$\begin{pmatrix} \beta_{1,g} \\ \beta_{2,gj} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} {\sigma_g}^2 & \sigma_{g,gj} \\ \sigma_{gj,g} & {\sigma_{gj}}^2 \end{pmatrix} \right)$$

Le terme $\sigma_{g,gj}$ représente la covariance entre les effets aléatoires, $\beta_{1,g}$ et $\beta_{2,gj}$.

Certains auteurs suggèrent de considérer les termes $(\alpha_{1,j} + \beta_{2,gj})$ comme la pente aléatoire spécifique à chaque groupe (Li, Hughes, Hemming, & Taljaard, 2021).

3.3.2 Closed cohort

Les essais *stepped wedge closed cohort* contrairement aux essais *cross-sectional* sont des essais dans lesquels les mesures sont répétées pour un même patient. Chaque patient est suivi à partir d'une période initiale et les mesures chez un même individu sont répétées dans le temps (Patterson, Leland, Mormer, & Palmerd, 2022). Un exemple d'utilisation de ce design pourrait être l'évaluation d'une nouvelle méthode pour accompagner les patients atteints de douleur chronique. Dans ce cas, le patient est évalué au cours de plusieurs périodes, des périodes pendant lesquelles la méthode standard est d'application et des périodes pendant lesquelles la nouvelle méthode est mise en place. En pratique, ce design est plus souvent utilisé comparé au design *cross-sectional*.

Une représentation à partir de la *figure 8* peut être réalisée afin de visualiser les essais *stepped wedge closed cohort*. Les deux premiers individus des deux premiers groupes $(g = 1, g = 2)$ peuvent être représentés comme suit :

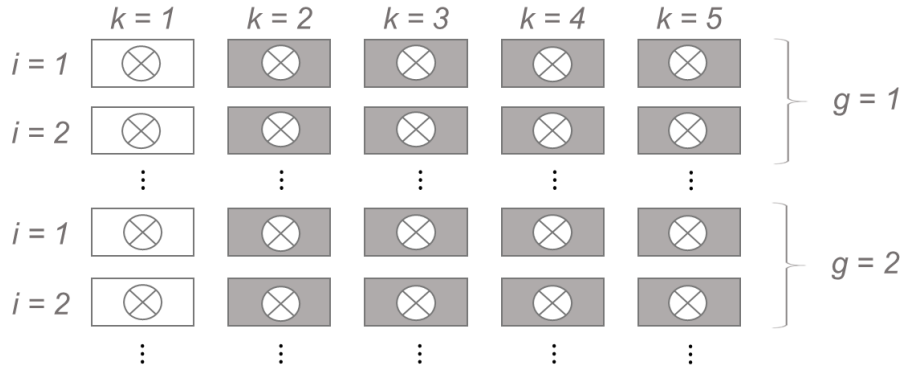

Figure 10 : Illustration des 2 premiers individus des deux premiers groupes dans le cas d'un essai stepped wedge closed cohort. Le traitement est représenté par un rectangle gris et le contrôle par un rectangle blanc. Les cibles représentent les mesures réalisées chez les patients.

Etant donné que plusieurs mesures sont prises chez le même individu *i*, sous les périodes de contrôle et de traitement, la variance intra-sujet est mesurable dans ce type d'essais *stepped wedge closed cohort*. Dans les différents modèles présentés dans cette section, il est considéré qu'une seule mesure par période pour chaque patient est récoltée.

3.3.2.1 Effet patient

Comme vu précédemment dans les essais croisés, la corrélation qui existe entre les différentes mesures récoltées chez un même patient peut être prise en compte (Li, Hughes, Hemming, & Taljaard, 2021).

Sur base du premier modèle proposé par Hussey et Hughes vu pour les essais *stepped wedge cross-sectional* (*8*), un effet patient est ajouté. L'observation du patient *i* appartenant au groupe *g* à la période *k* est alors modélisée comme :

$$y_{igkj} = \mu + \alpha_{1,j} treat_j + \alpha_{2,k}{}' p_k + \beta_{1,g} + \beta_{2,ig} + \epsilon_{igkj}$$

$$i = 1, \dots, n \quad g = 1, .., G \quad k = 1, \dots, K \quad j = 0,1$$

avec

$\beta_{2,ig}$: effet du patient *i* du groupe *g*

et la distribution des effets aléatoires est supposée normale :

$$\beta_{1,g} \sim N(0, {\sigma_g}^2)$$

$$\beta_{2,ig} \sim N(0, {\sigma_x}^2)$$

La variable période est considérée comme qualitative. La première période est considérée comme la période de référence. L'effet patient est, comme précédemment dans les essais croisés, considéré comme aléatoire ayant une distribution normale. $\hat{\beta}_{1,g}$ correspond à la différence du groupe *g* par rapport à la moyenne globale. $\hat{\beta}_{2,ig}$ représente la différence du

patient *i* par rapport à la moyenne globale. Les termes $\beta_{1,g}$ et $\beta_{2,ig}$ sont supposés indépendants. μ représente la moyenne de tous les groupes de tous les patients à la période de référence pour la situation de contrôle.

Ce modèle peut se réécrire sous forme matricielle $Y = X\alpha + Z\beta + \epsilon$ en considérant le premier patient du premier groupe et le premier patient du deuxième groupe à partir de la *figure 10*. Avec $X_{ig,1*1:2}$ et $Z_{ig,1*1:2}$ les designs matrices des effets fixes et des effets aléatoires, respectivement, pour le premier patient du premier groupe et celui du deuxième groupe, la forme matricielle est :

$$X_{ig,1*1:2} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix} \quad \alpha = \begin{pmatrix} \mu \\ \alpha_{1,1} \\ \alpha_{2,2} \\ \alpha_{2,3} \\ \alpha_{2,4} \\ \alpha_{2,5} \end{pmatrix}$$

et

$$Z_{ig,1*1:2} = \begin{pmatrix} 1 & 1 \cdots & 0 & 0 \cdots \\ 1 & 1 \cdots & 0 & 0 \cdots \\ 1 & 1 \cdots & 0 & 0 \cdots \\ 1 & 1 \cdots & 0 & 0 \cdots \\ 1 & 1 \cdots & 0 & 0 \cdots \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 \cdots & 1 & 1 \cdots \\ 0 & 0 \cdots & 1 & 1 \cdots \\ 0 & 0 \cdots & 1 & 1 \cdots \\ 0 & 0 \cdots & 1 & 1 \cdots \\ 0 & 0 \cdots & 1 & 1 \cdots \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_{1,1} \\ \beta_{2,11} \\ \cdots \\ \beta_{1,2} \\ \beta_{2,12} \\ \cdots \end{pmatrix}$$

Ce modèle engendre deux nouvelles corrélations. La première est la corrélation *'within-individual'* et notée ρ_{wi}, qui correspond à la corrélation entre deux observations provenant d'un même individu. La deuxième est la corrélation *'between-individual'* et notée ρ_{bi}, qui correspond à la corrélation entre deux observations provenant de différents individus d'un même groupe *g*.

$$\rho_{wi} = \frac{{\sigma_g}^2 + {\sigma_x}^2}{({\sigma_g}^2 + {\sigma_x}^2 + \sigma^2)}$$

$$\rho_{bi} = \frac{{\sigma_g}^2}{({\sigma_g}^2 + {\sigma_x}^2 + \sigma^2)}$$

Si l'effet patient est rajouté au modèle complet (*10*) considérant l'interaction groupe-période, ce dernier devient alors :

$$y_{igkj} = \mu + \alpha_{1,j} treat_j + \alpha_{2,k}{}' p_k + \beta_{1,g} + \beta_{2,gk} + \beta_{3,ig} + \epsilon_{igkj} \qquad (12)$$

Les trois termes aléatoires sont supposés indépendants. La forme matricielle de ce modèle $Y = X\alpha + Z\beta + \epsilon$ avec $X_{ig,1*1:2}$ et $Z_{ig,1*1:2}$ les designs matrices des effets fixes et des effets aléatoires, respectivement, pour le premier patient du premier groupe et celui du deuxième groupe est la suivante :

$$X_{ig,1*1:2} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix} \quad \alpha = \begin{pmatrix} \mu \\ \alpha_{1,1} \\ \alpha_{2,2} \\ \alpha_{2,3} \\ \alpha_{2,4} \\ \alpha_{2,5} \end{pmatrix}$$

et

$$Z_{ig,1*1:2} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 1 \cdots & 0 & 0 & 0 \ldots & 0 & 0 \cdots \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 \cdots & 0 & 0 & 0 \cdots & 0 & 0 \cdots \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 \cdots & 0 & 0 & 0 \cdots & 0 & 0 \cdots \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 \cdots & 0 & 0 & 0 \cdots & 0 & 0 \cdots \\ 1 & 0 & 0 & 0 & 0 & 1 & 1 \cdots & 0 & 0 & 0 \cdots & 0 & 0 \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \ldots & 1 & 1 & 0 \cdots & 0 & 1 \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \cdots & 1 & 0 & 1 \cdots & 0 & 1 \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \cdots & 1 & 0 & 0 \cdots & 0 & 1 \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \cdots & 1 & 0 & 0 \cdots & 0 & 1 \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \cdots & 1 & 0 & 0 \cdots & 1 & 1 \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_{1,1} \\ \beta_{2,11} \\ \beta_{2,12} \\ \beta_{2,13} \\ \beta_{2,14} \\ \beta_{2,15} \\ \beta_{3,11} \\ \cdots \\ \beta_{1,2} \\ \beta_{2,21} \\ \beta_{2,22} \\ \cdots \\ \beta_{2,25} \\ \beta_{3,12} \\ \cdots \end{pmatrix}$$

Ce modèle contient trois effets aléatoires. Ces derniers sont l'effet groupe, l'effet d'interaction groupe-période et l'effet patient, considérés comme indépendants les uns des autres. Les termes de corrélation de ce modèle sont les suivants :

$$\rho_{wi} = \frac{{\sigma_g}^2 + {\sigma_x}^2}{({\sigma_g}^2 + {\sigma_{gk}}^2 + {\sigma_x}^2 + \sigma^2)}$$

$$\rho_{wp} = \frac{{\sigma_g}^2 + {\sigma_{gk}}^2}{({\sigma_g}^2 + {\sigma_{gk}}^2 + {\sigma_x}^2 + \sigma^2)}$$

$$\rho_{bp} = \frac{{\sigma_g}^2}{({\sigma_g}^2 + {\sigma_{gk}}^2 + {\sigma_x}^2 + \sigma^2)}$$

Où ρ_{wi} représente la corrélation qui existe entre les mesures répétées d'un même patient *i*. Les corrélations ρ_{wp} et ρ_{bp} représentent, comme vu précédemment, la corrélation entre deux observations de patients différents dans un même groupe *g* à une même période *k* et à deux périodes différentes, respectivement.

Les auteurs envisagent un modèle similaire au modèle (*11*), vu dans le contexte des essais *stepped wedge cross-sectional* (point 3.3.1.3). Cependant dans le cas des *stepped wedge closed cohort,* il est nécessaire d'inclure une matrice de corrélation pour les observations provenant d'un même patient. Une possibilité pour cela est de rajouter au modèle (*11*) une matrice de structure autorégressive au terme d'erreur ϵ_{igkj} pour le *i* *ième* patient du groupe *g*. La distribution du terme d'erreur devient alors :

$$\epsilon_{igkj} \sim N(0, \boldsymbol{R})$$

La matrice **R** est une matrice en blocs par patient qui reprend donc la corrélation qui existe entre les mesures répétées chez un même patient comme décrit lors des analyses des essais croisés. De plus, la corrélation qui existe entre deux patients d'un même groupe est reprise dans la matrice **G** (voir (*11*)). Pour rappel, deux observations provenant de deux groupes différents sont considérées comme indépendantes.

3.3.2.2 Effet d'interaction traitement-groupe

Les auteurs proposent un modèle plus 'complet' reprenant en plus des paramètres vus précédemment, un effet aléatoire d'interaction traitement-groupe (Li, Hughes, Hemming, & Taljaard, 2021). Les auteurs ne considèrent plus les effets aléatoires comme étant indépendants ce qui rend ce modèle plus complexe.

Sur base de (*12*), un effet aléatoire d'interaction traitement-groupe est ajouté. Le modèle devient :

$$y_{igkj} = \mu + \alpha_{1,j} treat_j + \alpha_{2,k}{}' p_k + \beta_{1,gk} + \beta_{2,gj} treat_j + \beta_{3,ig} + \epsilon_{igkj}$$

$$i = 1, \dots, n \quad g = 1, .., G \quad k = 1, \dots, K \quad j = 0,1$$

Avec les effets fixes :

μ : terme d'intercept

$\alpha_{1,j}$: effet traitement

$\alpha_{2,k}$: effet période

Et les effets aléatoires :

$\beta_{1,gk}$: effet d'interaction groupe-période

$\beta_{2,gj}$: effet d'interaction traitement-groupe

$\beta_{3,ig}$: effet de l'individu *i* du groupe *g*

En considérant le premier patient du premier groupe et le premier du deuxième groupe (*figure 10*) et la première période comme la période de référence, la forme matricielle $Y = X\alpha + Z\beta + \epsilon$ de ce modèle est :

$$X_{ig,1*1:2} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix} \quad \alpha = \begin{pmatrix} \mu \\ \alpha_{1,1} \\ \alpha_{2,2} \\ \alpha_{2,3} \\ \alpha_{2,4} \\ \alpha_{2,5} \end{pmatrix}$$

et

$$Z_{ig,1*1:2} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 1 \cdots & 0 & 0 & \ldots & 0 & 0 & 0 \ldots \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 \cdots & 0 & 0 & \cdots & 0 & 0 & 0 \cdots \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \cdots & 0 & 0 & \cdots & 0 & 0 & 0 \cdots \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \cdots & 0 & 0 & \cdots & 0 & 0 & 0 \cdots \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 \cdots & 0 & 0 & \cdots & 0 & 0 & 0 \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \ldots & 1 & 0 & \cdots & 0 & 0 & 1 \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \ldots & 0 & 1 & \cdots & 0 & 1 & 1 \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \ldots & 0 & 0 & \cdots & 0 & 1 & 1 \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \cdots & 0 & 0 & \cdots & 0 & 1 & 1 \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \cdots & 0 & 0 & \cdots & 1 & 1 & 1 \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_{1,11} \\ \beta_{1,12} \\ \beta_{1,13} \\ \beta_{1,14} \\ \beta_{1,15} \\ \beta_{2,11} \\ \beta_{3,11} \\ \cdots \\ \beta_{1,21} \\ \beta_{1,22} \\ \cdots \\ \beta_{1,25} \\ \beta_{2,21} \\ \beta_{3,12} \\ \cdots \end{pmatrix}$$

Les auteurs considèrent que $\beta_{1,gk}$ et $\beta_{2,gj}$ ne sont pas indépendants. Étant donné que la variable période est considérée comme qualitative, des $\hat{\beta}_{1,gk}$ sont prédits pour chacune des périodes pour chacun des groupes. Sur base de *(11)*, les $\beta_{1,gk}$ et $\beta_{2,gj}$ sont supposés être normalement distribués :

$$(\beta_{1,gk=1}, \beta_{1,gk=2}, \dots, \beta_{1,gk=K}, \beta_{2,gj})' \sim N\left(\begin{pmatrix} 0\boldsymbol{M} \\ 0 \end{pmatrix}, \begin{pmatrix} \mathbf{G} & \sigma_{gk,gj}\boldsymbol{M} \\ \sigma_{gj,gk}\boldsymbol{M}' & {\sigma_{gj}}^2 \end{pmatrix}\right)$$

La matrice **M** est une matrice de dimension $K \times 1$ uniquement constituée de 1.

Dans ce modèle, **G** est une matrice de structure *Toeplitz* de dimension $K \times K$ comme la matrice G présentée au modèle (*11*). Pour rappel, la matrice $G_{g,1}$ pour un groupe, en considérant une matrice *Toeplitz*, se présentait comme :

$$G_{g,1} = \begin{pmatrix} {\sigma_{gk}}^2 & \sigma_1 & \sigma_2 & \sigma_3 & \sigma_4 \\ \sigma_1 & {\sigma_{gk}}^2 & \sigma_1 & \sigma_2 & \sigma_3 \\ \sigma_2 & \sigma_1 & {\sigma_{gk}}^2 & \sigma_1 & \sigma_2 \\ \sigma_3 & \sigma_2 & \sigma_1 & {\sigma_{gk}}^2 & \sigma_1 \\ \sigma_4 & \sigma_3 & \sigma_2 & \sigma_1 & {\sigma_{gk}}^2 \end{pmatrix}$$

En supposant une structure Toeplitz, le nombre de paramètres à estimer est de 5 (1 paramètre de variance, noté ${\sigma_{gk}}^2$ et 4 paramètres de covariances, notés σ_{kt}).

${\sigma_{gk}}^2$ est la covariance entre deux observations d'un même groupe à la même période ($k = k'$) et σ_{kt} est la covariance entre deux observations d'un même groupe à deux périodes différentes ($k \neq t$).

Les auteurs considèrent $\beta_{1,gk}$ et $\beta_{2,gj}$ comme étant indépendants de $\beta_{3,ig}$ et ϵ_{igkj}. Ces deux derniers termes sont considérés indépendants et de distribution normale :

$$\beta_{3,ig} \sim N(0, {\sigma_x}^2)$$

$$\epsilon_{igkj} \sim N(0, \sigma^2)$$

Sur base de ce modèle, différentes structures de corrélation sont envisagées par les auteurs. Plusieurs de ces structures vont être présentées dont une structure complexe montrant l'implication du paramètre $\sigma_{gk,gj}$ dans le calcul de la corrélation entre deux observations sous les conditions de traitement (Li, Hughes, Hemming, & Taljaard, 2021).

Pour deux observations d'un même groupe qui sont mesurées sous les conditions de contrôle, les corrélations *'within-individual'*, *'within-period'* et *'between-period'* deviennent :

$$\rho_{wi} = \frac{\sigma_{kt} + {\sigma_x}^2}{({\sigma_{gk}}^2 + {\sigma_x}^2 + \sigma^2)}$$

$$\rho_{wp} = \frac{{\sigma_{gk}}^2}{({\sigma_{gk}}^2 + {\sigma_x}^2 + \sigma^2)}$$

$$\rho_{bp} = \frac{\sigma_{kt}}{(\sigma_{gk}{}^2 + \sigma_x{}^2 + \sigma^2)}$$

Le terme de covariance apparaît dans le calcul de la corrélation entre deux observations d'un même groupe mesurées sous les conditions de traitement. Les corrélations deviennent alors :

$$\rho_{wi} = \frac{\sigma_{kt} + \sigma_{gj}{}^2 + 2\sigma_{gk,gj} + \sigma_x{}^2}{(\sigma_{gk}{}^2 + \sigma_x{}^2 + \sigma_{gj}{}^2 + 2\sigma_{gk,gj} + \sigma^2)}$$

$$\rho_{wp} = \frac{\sigma_{gk}{}^2 + \sigma_{gj}{}^2 + 2\sigma_{gk,gj}}{(\sigma_{gk}{}^2 + \sigma_x{}^2 + \sigma_{gj}{}^2 + 2\sigma_{gk,gj} + \sigma^2)}$$

$$\rho_{bp} = \frac{\sigma_{kt} + \sigma_{gj}{}^2 + 2\sigma_{gk,gj}}{(\sigma_{gk}{}^2 + \sigma_x{}^2 + \sigma_{gj}{}^2 + 2\sigma_{gk,gj} + \sigma^2)}$$

Les auteurs présentent une troisième possibilité. Cette dernière est la corrélation entre une mesure prise sous les conditions de contrôle et une mesure prise sous les conditions de traitement (du même groupe). Les corrélations intra-individu et inter-période deviennent alors :

$$\rho_{wi} = \frac{\sigma_{kt} + \sigma_{gk,gj} + \sigma_x{}^2}{\sqrt{(\sigma_x{}^2 + \sigma_{gk}{}^2 + \sigma^2)}\sqrt{(\sigma_x{}^2 + \sigma_{gk}{}^2 + \sigma_{gj}{}^2 + 2\sigma_{gk,gj} + \sigma^2)}}$$

$$\rho_{bp} = \frac{\sigma_{kt} + \sigma_{gk,gj}}{\sqrt{(\sigma_x{}^2 + \sigma_{gk}{}^2 + \sigma^2)}\sqrt{(\sigma_x{}^2 + \sigma_{gk}{}^2 + \sigma_{gj}{}^2 + 2\sigma_{gk,gj} + \sigma^2)}}$$

Dans ce dernier modèle, il aurait été possible de rajouter l'effet aléatoire groupe, $\beta_{4,g}$, afin d'avoir un modèle complet étant donné que ce dernier est impliqué dans des effets d'interaction. Les auteurs ne l'ont cependant pas rajouté. Il est vrai que l'ajout de cet effet aléatoire compliquerait d'autant plus le modèle et la structure de dépendance entre les effets aléatoires.

3.3.2.3 Modélisation de l'effet traitement

Comme expliqué précédemment, les essais *stepped wedge closed cohort* sont caractérisés par le suivi des mêmes patients dans différents groupes. Ces individus passent progressivement de la situation de contrôle à la situation de traitement. Différentes mesures pour chaque individu sont récoltées soit sous les conditions de contrôle soit sous les conditions de traitement en fonction des différentes périodes. Plusieurs périodes peuvent alors s'écouler entre le moment de la mise en place du traitement et le moment de la mesure. Il est donc légitime de penser que le temps pourrait avoir un impact différent sur l'effet du traitement en fonction du nombre de périodes écoulées depuis la mise en place du traitement. Ce paragraphe résume donc quelques méthodes afin de ne plus tenir compte d'un effet traitement constant, comme c'était le cas jusqu'à présent.

Différentes modélisations de l'effet traitement, noté $\alpha_{1,j}$, en tenant compte des périodes écoulées sont possibles (Li, Hughes, Hemming, & Taljaard, 2021). Deux méthodes permettant de tenir compte de cet effet vont être présentées. Pour modéliser cet effet, il faut tenir compte d'une notion évoquée dans l'introduction des essais *stepped wedge*, les séquences. Pour rappel, les séquences, notées *s*, représentent le nombre de passages de la situation de contrôle à la situation de traitement. Cette notion permet de tenir compte du nombre de périodes écoulées depuis la mise en place du traitement, différent entre les groupes. Une précision importante est à faire sur la manière d'indicer cet effet séquence. L'indice de l'effet séquence du groupe *g* correspond à la période à laquelle le traitement est mis en place dans ce groupe, différente en fonction du groupe auquel le patient appartient. Les illustrations de cette section sont présentées à partir de l'exemple de la *figure 8*, dans lequel le traitement est implémenté dans le premier groupe à la deuxième période. Dans cet exemple, tous les groupes sont soumis à une période de contrôle ($k = 1$), la séquence du premier groupe, recevant le traitement à la *2 ième* période, correspond alors à $s = 2$.

Une première méthode serait de considérer cet effet comme étant spécifique à chaque nombre de périodes écoulées depuis le traitement. L'effet traitement se décomposerait alors comme suit :

$$\alpha_{1,j}{}' = \alpha_{1,j(1|k=s)} + \alpha_{1,j(2|k=s+1)} + \cdots + \alpha_{1,j(K-1|k=K)}$$

$$s = 2, \ldots, K \quad k = s, \ldots, K$$

Les auteurs appellent cette représentation 'une représentation stationnaire' (Li, Hughes, Hemming, & Taljaard, 2021). Cette appellation vient du fait qu'un estimateur $\hat{\alpha}_{1,j}$ est obtenu uniquement pour les observations où $k \geq s$, c'est-à-dire à la période de la mise en place du traitement et les périodes qui se sont écoulées depuis cette mise en place. L'estimation de la situation de contrôle sera incluse dans le terme d'intercept, comme précédemment.

Ci-dessous un schéma a été réalisé afin de mieux comprendre la construction de ce paramètre (Li, Hughes, Hemming, & Taljaard, 2021), ce schéma est réalisé à partir de l'illustration de la *figure 8*. Pour faciliter les notations du schéma ci-dessous, $\alpha_{1,j(1|k=s)}$ est noté α_1 , $\alpha_{1,j(2|k=s+1)}$ est noté α_2 etc…

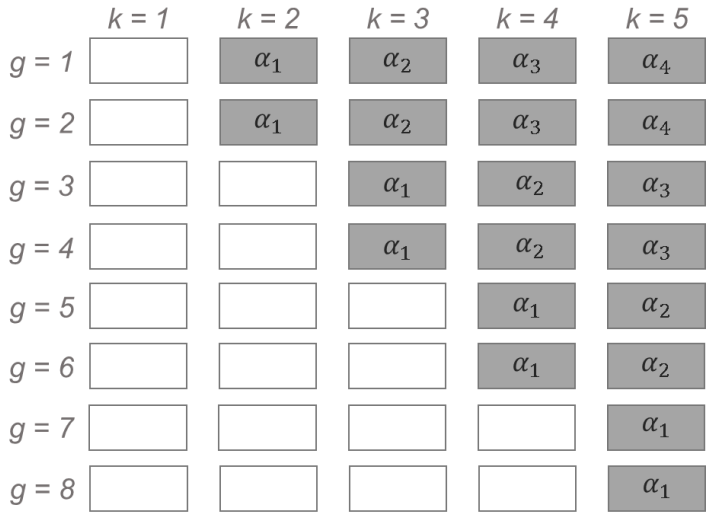

Figure 11 : Illustration de la 'représentation stationnaire' de l'effet traitement dans les essais stepped wedge. Les carrés blancs correspondent à la situation de contrôle tandis que les carrés gris représentent le traitement.

Avec cette première méthode, $K-1$ paramètres pour l'effet traitement en tenant compte du nombre de périodes écoulées depuis la mise en place de ce dernier sont estimés. Cependant, cette méthode présente un inconvénient. En effet, le nombre de paramètres à estimer peut vite devenir important en fonction du nombre de périodes de l'essai. Afin de contrer cet inconvénient, les auteurs proposent une deuxième méthode. L'idée est alors d'inclure un effet traitement et de rajouter un terme qui tient compte du temps écoulé, comme par exemple :

$$\alpha_{1,j} = \alpha_{1,j\,(0|k=s)} + \alpha_{1,j\,(1|k>s)}(k-s) \qquad s=2,\dots,K \;\; k=s,\dots,K$$

Le paramètre $\alpha_{1,j(0)}$ correspond à l'effet au moment où le traitement est mis en place. Le paramètre $\alpha_{1,j(1)}$ représente, quant-à-lui, l'effet de ce traitement depuis sa mise en place. Ce terme sera donc multiplié par $(k-s)$, le nombre de périodes *k* écoulées depuis la mise en place du traitement. L'estimation de la situation de contrôle sera incluse dans le terme d'intercept.

Ci-dessous un second schéma a été réalisé (Li, Hughes, Hemming, & Taljaard, 2021) pour simplifier le schéma $\alpha_{1,j\,(0)}$ est noté α_0 et $\alpha_{1,j\,(1)}$ est noté α_1:

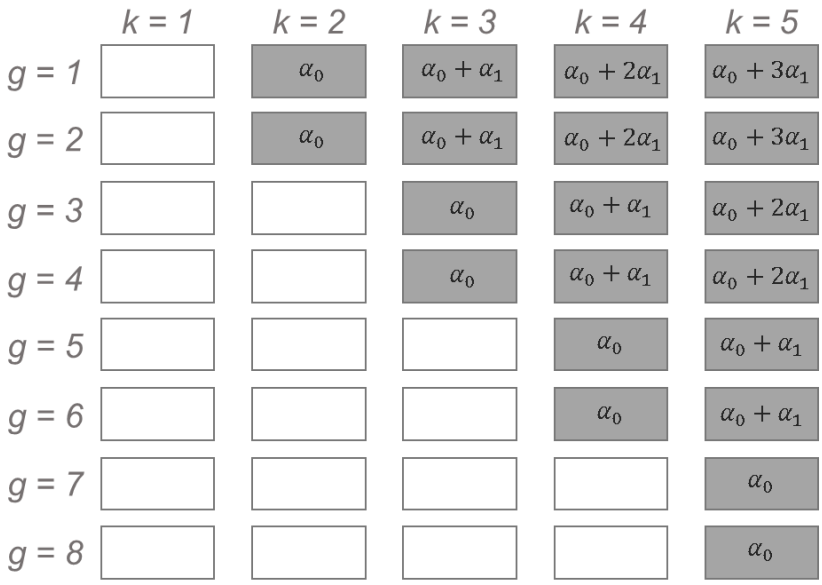

Figure 12 : Illustration de l'effet traitement linéaire en fonction du temps des essais stepped wedge. Les carrés blancs correspondent à la situation de contrôle tandis que les carrés gris représentent le traitement.

Le paramètre α_j peut être considéré comme l'effet traitement constant plus un terme d'interaction linéaire en fonction du temps. Cette deuxième méthode a comme hypothèse que les périodes sont espacées de manière équivalente.

3.4 Analyses statistiques des essais *N of 1*

Pour rappel, les essais *N of 1* sont des essais dans lesquels les patients reçoivent à plusieurs reprises deux traitements, ou plus, à des moments distincts (Senn S. , 2019). Pour décrire les techniques d'analyse de ces essais, les traitements seront représentés par les lettres A et B, comme dans la description des analyses statistiques des essais croisés vus précédemment. Dans les analyses des essais *N of 1*, la randomisation des traitements va être considérée comme étant une randomisation par bloc (voir point 2.3). Pour rappel, une randomisation par bloc signifie que les deux traitements sont assignés au hasard par bloc de deux, dans le cas où deux traitements sont étudiés. Chaque bloc est composé de deux périodes au cours desquelles le patient reçoit soit le traitement A puis le traitement B, soit le traitement B puis le A. Un exemple d'une structure des essais *N of 1* est présenté à la figure ci-dessous :

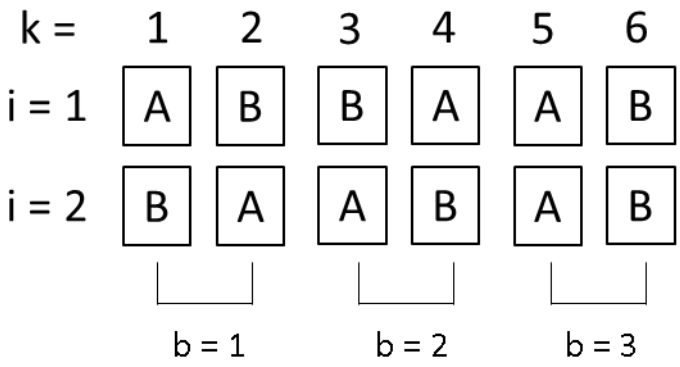

Figure 13 : Structure d'un essai N of 1 (exemple fictif), comparant deux traitements A et B, réalisé sur deux patients (i =1,2) avec k représentant les périodes (k=1…6) et b représentant les blocs (b=1,2,3).

La *figure 13*, ci-dessus, représente donc un essai *N of 1* sur un seul patient. Deux traitements, A et B, sont administrés au patient durant plusieurs périodes, notées *k*. L'essai repris dans cette illustration est constitué de 3 blocs, chacun contenant 2 périodes donc un total de 6 périodes. La randomisation a donc été appliquée pour chacun des blocs afin de savoir quel traitement est administré au cours de la première période du bloc.

Les essais *N of 1* ne sont pas des essais qui ont été largement utilisés (Chen & Chen, 2014). Comme vu précédemment (point 2.3), l'utilisation de ce type d'essais n'est permise que lorsque certaines conditions sont réunies (par exemple, des maladies stables dans le temps, des traitements à effet rapide et court …). De plus, l'analyse de ce type d'essais est relativement complexe. Pour rappel, les essais *N of 1* avaient comme premier objectif d'étudier l'efficacité de différents traitements sur un individu spécifique. Cependant, depuis quelques années, certains auteurs envisagent d'utiliser les essais *N of 1* pour étudier l'effet traitement sur plusieurs individus. Bien que le nombre d'individus repris dans des études *N of 1* soit rarement très élevé au vu des conditions de ce type d'essais. Il est également important de considérer ici qu'étant donné que l'administration de plusieurs traitements est répétée un certain nombre de fois, une forte corrélation entre les observations doit être considérée dans les analyses.

Plusieurs auteurs ont proposé des méthodes différentes utilisant des modèles mixtes (Chen & Chen, 2014). Ces auteurs présentent généralement des modèles pour étudier l'effet traitement sur plusieurs individus. Quelques-uns d'entre eux envisagent, cependant, des modèles lorsque l'essai ne contient qu'un seul patient. Certaines de ces méthodes vont être présentées et résumées afin de comprendre la particularité et la difficulté de l'analyse des essais *N of 1*.

Une remarque importante pour la suite des analyses est que la plupart des auteurs considèrent des modèles sans intercept avec des effets aléatoires non centrés en zéro. Pour simplifier la compréhension et rester dans la continuité par rapport à la présentation des modèles d'analyse des deux designs précédents, les effets aléatoires des modèles présentés pour les essais *N of 1* seront, pour ce travail, centrés en zéro. Les termes fixes correspondant à leur moyenne seront ajoutés dans le modèle. De plus, la majorité des auteurs ayant rédigé des articles sur les essais *N of 1* considère dans leurs analyses que les patients ont reçu le même nombre de fois les deux traitements et qu'une seule mesure est réalisée par période.

3.4.1 Modèles de base

Parmi les auteurs qui se sont intéressés à l'analyse des essais *N of 1,* deux stratégies principales se dégagent dans la littérature en ce qui concerne le choix de la variable à modéliser. Une première partie des auteurs utilise directement la différence entre les traitements comme variable réponse. Tandis que la seconde partie des auteurs utilise l'observation y_{ik} réalisée pour le patient *i* à la période *k* comme précédemment, c'est-à-dire la mesure réalisée après l'administration d'un traitement à une certaine période.

En plus de cette particularité, plusieurs auteurs proposent de considérer l'effet traitement comme étant un effet aléatoire dans l'analyse des essais *N of 1*. Jusqu'à présent dans les essais croisés, l'effet traitement était considéré comme un effet fixe. En considérant l'effet traitement pour un patient comme étant une réalisation (aléatoire) de l'effet traitement, cela permet d'étudier l'effet traitement propre à chaque patient. Mais surtout, cet effet permettrait d'obtenir une estimation de la variabilité de l'effet traitement entre les patients ou au sein d'un patient.

Dans cette section, les deux modèles principaux de la littérature vont être présentés.

3.4.1.1 Effet traitement fixe

Ce premier modèle, présenté par les auteurs Chen et Chen (Chen & Chen, 2014), considère la variable Y comme étant la différence entre les valeurs obtenues pour chaque traitement des deux périodes de chaque bloc. Ce n'est plus la mesure qui est modélisée ici, comme vu précédemment mais bien une différence. Dans ce premier modèle, l'effet traitement est donc indirectement inclus dans le modèle grâce à la variable Y, la différence entre les valeurs des deux traitements de chaque bloc.

Une simple modélisation de la différence entre les deux traitements du bloc *b* du patient *i* est :

$$y_{ib} = \mu + \beta_{1,i} + \epsilon_{ib}$$
$$i = 1, \dots, n \quad b = 1, \dots, B$$

avec

μ : terme d'intercept

$\beta_{1,i}$: effet patient

Les termes $\beta_{1,i}$ et ϵ_{ib} sont supposés suivre une distribution normale :

$$\beta_{1,i} \sim N(0, {\sigma_x}^2)$$
$$\epsilon_{ib} \sim N(0, \sigma^2)$$

Ce premier modèle prend en compte le fait qu'il existe un potentiel effet patient, c'est-à-dire que la différence entre les traitements à chaque bloc peut être spécifique et propre à chaque patient. Un $\hat{\beta}_{1,i}$ peut être prédit pour chacun des patients. Il représente la déviation de l'intercept de l'individu *i* par rapport à la moyenne des différences de tous les blocs de tous les patients. Dans ce modèle, μ représente la moyenne des différences des traitements de tous les blocs de tous les patients. ϵ_{ib} représente l'erreur du sujet *i* dans le bloc *b*.

La composante de variance, ${\sigma_x}^2$, représente la variance de l'effet traitement entre les patients et σ^2 la variance entre les mesures répétées par bloc par patient. Étant donné que y_{ib} est la différence des traitements du bloc *b*, certains auteurs définissent la variance du terme d'erreur comme étant 2 fois la variance intra-patient, notée ${\sigma_w}^2$ (Senn S. , 2019). Alors, la distribution du terme ϵ_{ib} est supposée être une normale :

$$\epsilon_{ib} \sim N(0, 2{\sigma_w}^2)$$

Les termes $\beta_{1,i}$ et ϵ_{ib} sont considérés comme étant indépendants, comme précédemment.

A partir de ce premier modèle, un nouveau modèle peut être construit. Ce dernier ne modélise plus la différence par bloc *b* mais la moyenne des différences des traitements pour un patient donné *i*. Cette dernière est estimée comme :

$$\bar{y}_{i.} = \frac{\sum_{b=1}^{B} y_{ib}}{B}$$

Avec y_{ib} qui représente donc la différence entre le traitement A et le traitement B du patient *i* au bloc *b* et B le nombre de blocs de l'étude. La modélisation de cette moyenne des différences obtenues par patient, notée $\bar{y}_{i.}$ est alors :

$$\bar{y}_{i.} = \mu + \beta_{1,i} + \epsilon_{i.} \qquad (13)$$

avec $\epsilon_{i.} = \sum_{b=1}^{B} \epsilon_{ib}/B$

Dans ce cas, la variance de $\bar{y}_{i.}$ est telle que (Araujo, Julious, & Senn, 2016):

$$Var(\bar{y}_{i.}) = {\sigma_x}^2 + ({2\sigma_w}^2/B)$$

L'estimation de la moyenne des différences de traitements de tous les patients peut être obtenue comme suit :

$$\bar{Y} = \frac{\sum_{i=1}^{n} \sum_{b=1}^{B} y_{ib}}{Bn}$$

La variance de cet estimateur sera alors :

$$Var(\bar{Y}) = \frac{{\sigma_x}^2 + (2{\sigma_w}^2/B)}{n}$$

Ces notions seront utilisées plus tard lors de la présentation du calcul de la taille d'échantillon des essais *N of 1* (voir point 4.3).

Étant donné que les essais *N of 1* sont répartis en plusieurs blocs au cours du temps, il est important de considérer le fait qu'un effet bloc pourrait impacter la différence des traitements. Les auteurs considèrent donc un modèle contenant un effet fixe bloc ainsi qu'un effet aléatoire patient (Chen & Chen, 2014).

Un effet fixe bloc est donc rajouté. La modélisation de y_{ib}, la différence entre les deux traitements du bloc *b* du patient *i*, devient :

$$y_{ib} = \mu + \alpha_{1,b} w_b + \beta_{1,i} + \epsilon_{ib}$$
$$i = 1, \dots, n \quad b = 1, \dots, B$$

avec

$\alpha_{1,b}$: effet bloc

$\beta_{1,i}$: effet du patient *i*

L'effet aléatoire et le terme d'erreur sont supposés suivre une distribution normale :

$$\beta_{1,i} \sim N(0, {\sigma_x}^2)$$
$$\epsilon_{ib} \sim N(0, \sigma^2)$$

La variable bloc est considérée comme une variable quantitative. Dans ce cas, w_b prend la valeur du bloc auquel correspond la différence des traitements ($w_b = 1$, lorsque y_{ib} correspond à la différence des traitements du bloc 1, $w_b = 2$, lorsque y_{ib} correspond au bloc 2,…). Le fait de considérer la variable bloc comme une variable quantitative suppose une relation linéaire entre la variable réponse et la variable bloc. Le paramètre $\alpha_{1,b}$ représente la pente de la droite de régression. Dans ce modèle, μ représente la moyenne des différences des traitements de tous les blocs de tous les patients.

Comme dans le modèle précédent, $\beta_{1,i}$ est l'effet aléatoire patient donc un $\hat{\beta}_{1,i}$ peut être prédit pour chacun des patients. Il représente la déviation de l'intercept de l'individu *i* par rapport à la moyenne de tous les blocs de tous les patients. Le terme ϵ_{ib} représente l'erreur du sujet *i* dans le bloc *b*. Comme précédemment, la composante de variance ${\sigma_x}^2$ correspond

à la variance entre les patients et σ^2 à la variance entre les mesures répétées par bloc pour chaque patient.

La forme matricielle $Y = X\alpha + Z\beta + \epsilon$ peut être construite à partir de ce modèle. Puisque dans ce modèle la différence entre les traitements par bloc pour chaque patient est modélisée, la matrice Y complète est de dimension $(nB \times 1)$. Dans le cas où la forme matricielle est représentée pour les 2 patients illustrés à la *figure 13*, alors la matrice Y est de dimension (6×1). A partir de la *figure 13 et* avec $X_{i,1:2}$ et $Z_{i,1:2}$ les designs matrices des effets fixes et aléatoires du premier et du deuxième patient, respectivement, la forme matricielle s'écrit comme suit :

$$X_{i,1:2} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} \quad \alpha = (\mu, \alpha_{1,b})' \quad \textbf{et} \quad Z_{i,1:2} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \quad \beta = (\beta_{1,1}, \beta_{1,2})'$$

Dans ce modèle, l'intercept représente donc la moyenne des différences de tous les blocs de tous les patients. C'est donc ce terme qui permet de détecter un effet traitement. En effet, lorsque l'intercept est différent de zéro cela signifie que la moyenne des différences entre les traitements est différente de zéro. Dans ce cas, les deux traitements ne sont pas similaires et il y a donc bien un effet traitement différent de zéro.

Les auteurs précisent que si l'essai n'est réalisé que sur un seul patient, alors le modèle devient :

$$y_b = \mu + \epsilon_b$$

ϵ_b représentera l'erreur aléatoire du bloc *b*.

Dans ce second modèle, l'intercept μ représente la moyenne des différences de tous les blocs pour le patient. Si l'intercept est différent de zéro alors il y a bien une différence entre les deux traitements.

Le nombre d'observations récoltées pour un essai *N of 1* réalisé sur un seul individu peut être très faible. En effet, lorsque l'étude ne comporte pas un nombre élevé de blocs alors le nombre d'observations à modéliser donc le nombre de différences entre les traitements par bloc est faible. Par conséquent, les modèles réalisés pour analyser les données d'un seul individu contiennent très peu de paramètres. Bien entendu comme le nombre d'observations disponibles dépend directement du nombre de blocs, il est possible de réaliser une étude sur une très longue période donc beaucoup de blocs. Cependant, en pratique au vu de l'objectif de l'étude, déterminer le meilleur traitement pour un patient spécifique, et des risques liés à l'attribution de plusieurs traitements successivement à un patient, le nombre de blocs est rarement élevé dans les études *N of 1*.

3.4.1.2 *Effet traitement aléatoire*

Dans ce deuxième paragraphe, un effet traitement est de nouveau introduit dans le modèle mais cette fois, cet effet est considéré comme aléatoire. Ce modèle a été proposé par les auteurs Araujo, Julious et Senn (Araujo, Julious, & Senn, 2016) sur base de l'article de Chen et Chen (Chen & Chen, 2014). Contrairement aux modèles présentés à la section précédente (section 3.4.1.1), les auteurs considèrent ici que la variable Y modélisée correspond à la mesure réalisée sur le patient *i* à la période *k* et non plus à la différence des traitements.

La modélisation de y_{ijkb}, la mesure réalisée sur le patient *i* suite à l'administration du traitement *j* à la période *k* du bloc *b,* devient alors :

$$y_{ijkb} = \mu + \alpha_{1,j} treat_j + \beta_{1,i} + \beta_{2,bi} + \beta_{3,ji} Z_{ikb} + \epsilon_{ijkb}$$
$$i = 1, \dots, n \quad j = A, B \quad k = 1, \dots, K \quad b = 1, \dots, B$$

avec

$\alpha_{1,j}$: effet traitement

$\beta_{1,i}$: effet du patient *i*

$\beta_{2,bi}$: effet bloc pour le patient *i*

$\beta_{3,ji}$: effet traitement pour le patient *i*

et

$$\beta_{1,i} \sim N(0, {\sigma_x}^2)$$
$$\beta_{2,bi} \sim N(0, {\sigma_b}^2)$$
$$\beta_{3,ji} \sim N(0, {\sigma_j}^2)$$
$$\epsilon_{ijkb} \sim N(0, \sigma^2)$$

$\beta_{1,i}$ est l'effet aléatoire patient donc $\hat{\beta}_{1,i}$ correspond à la déviation de l'intercept du patient *i* par rapport à la moyenne globale. Les auteurs considèrent la variable bloc qui permet de tenir compte du temps dans le modèle comme une variable qualitative. $\beta_{2,bi}$ représente la déviation du bloc *b* du patient *i* par rapport à la moyenne globale. $\alpha_{1,j}$ correspond à l'effet traitement. $treat_j$ est une variable indicatrice, $treat_j$ prendra la valeur de 0 si le patient est soumis au traitement de référence et 1 sinon. $\beta_{3,ji}$ représente l'effet traitement pour le patient *i*. Z_{ikb} est une variable indicatrice. Cette dernière prendra la valeur de $1/2$ si le patient a reçu le traitement A à la période *k* du bloc *b* et $-1/2$ si c'est le traitement B. Ce système de codage de l'effet traitement est utilisé pour étudier un effet traitement aléatoire, donc la différence entre les traitements, spécifique au patient *i*. Dans ce modèle, le terme d'intercept μ représente en réalité la moyenne globale de tous les patients pour tous les blocs pour le traitement de référence.

Les termes aléatoires sont considérés indépendants.

Ce modèle peut être réécrit sous la forme matricielle : $Y = X\alpha + Z\beta + \epsilon$. En partant de la *figure 13* et avec $X_{i,1:2}$ et $Z_{i,1:2}$ les designs matrices des effets fixes et aléatoires de ces patients, respectivement, et A pris comme traitement de référence, la forme matricielle s'écrit comme suit :

$$X_{i,1:2} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 1 \\ 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 1 & 1 \\ 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 1 & 0 \\ 1 & 1 \end{pmatrix} \quad \alpha = (\mu, \alpha_{1,B})'$$

et

$$Z_{i,1:2} = \begin{pmatrix} 1 & 1 & 0 & 0 & 1/2 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & -1/2 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & -1/2 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1/2 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1/2 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & -1/2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & -1/2 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & -1/2 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1/2 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & -1/2 \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_{1,1} \\ \beta_{2,11} \\ \beta_{2,21} \\ \beta_{2,31} \\ \beta_{3,j1} \\ \beta_{1,2} \\ \beta_{2,12} \\ \beta_{2,22} \\ \beta_{2,32} \\ \beta_{3,j2} \end{pmatrix}$$

Plusieurs composantes de variance découlent de ce modèle. Comme précédemment, ${\sigma_x}^2$ représente la variance entre les patients. ${\sigma_j}^2$ est la variance de l'effet traitement entre les patients. ${\sigma_b}^2$ est la variance de l'effet bloc entre les patients. Enfin, σ^2 représente la variance entre les observations répétées par bloc chez un même patient.

Dans ce modèle, l'effet traitement est estimé grâce au paramètre $\alpha_{1,j}$. Lorsque la variance entre les traitements ${\sigma_j}^2$ est égale à 0, l'effet traitement entre les patients est identique. En effet, lorsque ${\sigma_j}^2 = 0$, l'effet traitement ne varie pas d'un patient à l'autre.

D'autres auteurs ont également travaillé sur base d'un modèle contenant un effet patient et un effet traitement, tous deux considérés comme des effets aléatoires (Krone, et al., 2020) (Zucker, Ruthazer, & Schmid, 2010). A la différence du modèle présenté ci-dessus, les auteurs ne considèrent pas un effet aléatoire bloc dans leur modèle.

3.4.2 Modèle à Covariance pattern

Certains auteurs mettent en place des modèles à covariance pattern pour l'analyse des essais *N of 1* (Zucker, Ruthazer, & Schmid, 2010). A partir de leur présentation de différents modèles et de différentes structures de corrélation, un modèle simple à covariance pattern va être construit.

La modélisation de y_{ijk}, la mesure réalisée sur la patient *i* après l'administration du traitement *j* à la période *k*, est :

$$y_{ijk} = \mu + \alpha_{1,j} treat_j + \epsilon_{ijk} \qquad i = 1, \dots, n \quad j = A, B$$

avec

$\alpha_{1,j}$: effet traitement

Puisque dans ce modèle l'effet traitement est considéré comme fixe, son effet sera estimé en fonction d'un traitement défini comme la référence, comme vu précédemment. Dans le cas où le traitement A est défini comme le traitement de référence, $treat_j$ prend la valeur de 0 lorsque l'observation est réalisée après l'administration du traitement A et la valeur de 1 si B. Le paramètre $\alpha_{1,j}$ représente la différence entre le traitement A et le traitement B. Le terme d'intercept μ représente la valeur de la moyenne de Y de tous les patients pour le traitement de référence, le traitement A ici.

Les termes d'erreur sont supposés normalement distribués :

$$\epsilon_{ijk} \sim N(0, \boldsymbol{R})$$

Comme vu précédemment dans les essais croisés (voir point 3.2.3.2), la matrice **R** permet de tenir compte de la corrélation qui existe entre les mesures répétées chez un même individu. Cette matrice aura une dimension de $K * K$ par patient, avec K qui correspond au nombre de périodes de l'étude. Cette matrice peut être structurée de différentes façons.

Les auteurs considèrent d'abord une matrice non structurée. Cependant, comme détaillé dans les sections précédentes, le nombre de paramètres à estimer d'une matrice non structurée peut vite devenir important en fonction du nombre de périodes. Dans un deuxième temps, les auteurs proposent une corrélation autorégressive. Pour rappel, cette structure correspond à une corrélation qui diminue au fur et à mesure que le temps entre les mesures augmente. Dans ce cas, le nombre de paramètres à estimer est fortement réduit. En effet, dans une structure autorégressive, seulement deux paramètres sont à estimer pour construire la matrice **R** (voir *formule 1*).

Cependant, il est important de rappeler que l'utilisation des modèles à covariance pattern peut engendrer l'estimation de nombreux paramètres. Ces modèles sont donc principalement utilisés dans les essais *N of 1* constitués de plusieurs individus.

De plus, les auteurs proposent également de mettre en place des matrice **R** en considérant les observations non corrélées (Zucker, Ruthazer, & Schmid, 2010). Cette structure implique que

la covariance entre les données provenant d'un même patient est égale à 0. Mais dans ce cas, ces modèles ne sont plus considérés comme des modèles à covariance pattern.

3.4.3 Effet *carry over*

Pour terminer les analyses des essais *N of 1*, un modèle permettant de mesurer et tenir compte de l'effet *carry over* dans l'étude de l'effet traitement va être présenté. Pour rappel, un effet *carry over* est un phénomène qui peut apparaître lorsque les effets d'un traitement sont prolongés dans le temps. La mesure réalisée n'est alors plus la mesure de l'effet d'un traitement mais bien des effets de plusieurs traitements cumulés. Étant donné que les essais *N of 1* sont caractérisés par l'administration de plusieurs traitements successivement, ils sont soumis au risque d'un effet *carry over* comme les essais croisés. Certains auteurs se sont donc concentrés sur l'estimation de cet effet (Chen & Chen, 2014). Afin d'estimer l'effet *carry over*, les auteurs ne vont plus s'intéresser à l'effet par bloc mais bien à l'effet par période. Pour rappel, dans les essais *N of 1* de deux traitements A et B, les traitements sont administrés par blocs, constitués chacun de deux périodes (voir *figure 13*).

Les auteurs proposent un premier modèle qui est utilisé dans le cas où l'essai *N of 1* est constitué de plusieurs individus.

La mesure du patient *i* à la période *k* ayant reçu le traitement *j* est modélisée comme suit :

$$y_{ijk} = \mu + \alpha_{1,j} treat_j + \alpha_{2,k} p_k + \alpha_{3,A} c_A + \alpha_{4,B} c_B + \beta_{1,i} + \epsilon_{ijk}$$

$$i = 1, \dots, n \quad k = 1, \dots, K \quad j = A, B$$

avec

$\alpha_{1,j}$: effet traitement

$\alpha_{2,k}$: effet période

$\alpha_{3,A}$: effet *carry over* du traitement A

$\alpha_{4,B}$: effet *carry over* du traitement B

$\beta_{1,i}$: effet du patient *i*

et les termes $\beta_{1,i}$ et ϵ_{ijk} sont supposés normalement distribués :

$$\beta_{1,i} \sim N(0, {\sigma_x}^2)$$

$$\epsilon_{ijk} \sim N(0, \sigma^2)$$

La variable $treat_j$ est une dummy variable. En considérant que A est le traitement de référence, si la mesure est réalisée après l'administration du traitement A alors $treat_j$ sera égal à 0 et à 1 si B. Le paramètre α_j représente donc la différence qui existe entre les deux traitements. Dans ce modèle, la variable période est considérée comme une variable quantitative. Le paramètre $\alpha_{2,k}$ correspond donc à l'effet période et p_k sera égale à la période à laquelle l'observation a été obtenue. $\alpha_{2,k}$ représente la pente de la droite de régression. L'effet patient étant un effet aléatoire, une prédiction $\hat{\beta}_{1,i}$ peut être obtenue pour chacun des

patients. $\hat{\beta}_{1,i}$ représente la déviation de l'intercept du patient *i* par rapport à la moyenne de tous les patients.

En ce qui concerne l'estimation de l'effet *carry over*, trois situations sont envisageables : un effet *carry over* provoqué par le traitement A, un effet *carry over* provoqué par le traitement B ou aucun effet *carry over*. c_A et c_B sont des dummy variable. c_A prendra la valeur de 1 lorsque le patient à la précédente période a reçu le traitement A et 0 sinon. c_B prendra la valeur de 1 lorsque le patient à la précédente période a reçu le traitement B et 0 sinon. Lorsque c_A et c_B sont tous les deux de 0 alors il n'y a pas d'effet *carry over* pour l'observation en question. Ce dernier cas s'observe pour la première observation récoltée à la première période qui n'est précédée par l'administration d'aucun traitement. Dans ce cas, le terme d'intercept μ correspond à une moyenne globale des valeurs du traitement A (lorsque A est le traitement de référence) qui ne sont précédées d'aucun traitement. Les paramètres $\alpha_{3,A}$ et $\alpha_{4,B}$ représentent les effets *carry over* des traitements A et B, respectivement. Les auteurs supposent que l'effet *carry over*, s'il est présent, provient uniquement de la période antérieure et non des autres périodes qui précèdent (Chen & Chen, 2014). $\alpha_{3,A}$ et $\alpha_{4,B}$ correspondent donc à l'effet laissé par les traitements A ou B, administrés à la période précédente.

La forme matricielle de ce modèle $Y = X\alpha + Z\beta + \epsilon$, en considérant les deux patients de la *figure 13* et avec $X_{i,1:2}$ et $Z_{i,1:2}$ les designs matrices des effets fixes et aléatoires de ces patients, respectivement est :

$$X_{i,1:2} = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 2 & 1 & 0 \\ 1 & 1 & 3 & 0 & 1 \\ 1 & 0 & 4 & 0 & 1 \\ 1 & 0 & 5 & 1 & 0 \\ 1 & 1 & 6 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 2 & 0 & 1 \\ 1 & 0 & 3 & 1 & 0 \\ 1 & 1 & 4 & 1 & 0 \\ 1 & 0 & 5 & 0 & 1 \\ 1 & 1 & 6 & 1 & 0 \end{pmatrix} \quad \alpha = \begin{pmatrix} \mu \\ \alpha_{1,B} \\ \alpha_{2,k} \\ \alpha_{3,A} \\ \alpha_{4,B} \end{pmatrix} \textbf{ et } \quad Z_{i,1:2} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_{1,1} \\ \beta_{1,2} \end{pmatrix}$$

Les auteurs présentent également un modèle dans le cas où un seul patient est présent dans l'étude. Le modèle pour le patient ayant reçu le traitement *j* à la période *k* devient :

$$y_{jk} = \mu + \alpha_{1,j} treat_j + \alpha_{2,A} c_A + \alpha_{3,B} c_B + \epsilon_{jk}$$

Dans ce modèle, ϵ_{jk} représente l'erreur aléatoire de la période *k*.

Le paramètre $\alpha_{1,j}$ correspond à l'effet traitement, comme dans le modèle précédent. Un effet période n'a pas été considéré par les auteurs. Au vu du nombre peu élevé d'observations pour un essai *N of 1* réalisé sur un seul patient, les auteurs considèrent un modèle contenant un nombre de paramètres minimum.

4 Taille d'échantillon

À présent que les analyses statistiques pour chacun des designs ont été développées en profondeur, il est important de se concentrer sur un point essentiel à l'analyse, le calcul de la taille d'échantillon. Le calcul de la taille d'échantillon est l'une des premières étapes d'un essai et est primordial puisqu'il permettra de répondre à la question de recherche tout en contrôlant le risque d'erreur de type I et de type II (Gupta, Attri, Singh, Kaur, & Kaur, 2016). Ce calcul représente l'équilibre entre l'acceptation d'un point de vue éthique et la mise en évidence statistique d'un effet. De fait, une taille d'échantillon trop petite ne permettrait pas de mettre en évidence un effet traitement tandis qu'une taille d'échantillon trop grande exposerait trop d'individus au risque du traitement.

Avant de s'engager dans les calculs de la taille d'échantillon, il est nécessaire de rappeler certains concepts importants. Quatre paramètres de base sont nécessaires au calcul de la taille d'échantillon (Gupta, Attri, Singh, Kaur, & Kaur, 2016).

Le premier est le niveau de significativité, noté α. Ce terme correspond au pourcentage de risque accepté de détecter une différence significative entre deux traitements alors qu'en réalité il n'existe aucune différence. Ce risque d'erreur est également connu sous le nom d'erreur de type I. Plus la taille de l'échantillon augmente (tout le reste étant constant), plus la probabilité de faire une erreur de type I diminue.

Un deuxième paramètre important est le concept de *puissance*. La puissance correspond à la probabilité de conclure, à l'aide de tests d'hypothèses avec un niveau de significativité choisi, que les effets ne sont pas égaux, en sachant qu'il existe une réelle différence entre les traitements. Cette puissance est directement liée à un deuxième type d'erreur, appelé erreur de type II. Cette erreur est définie comme la probabilité de ne pas mettre en évidence une différence d'effet alors que cette différence existe belle et bien. Elle est notée β. La puissance est alors calculée comme $1 - \beta$. Plus la puissance d'un test augmente (tout le reste étant constant), plus la probabilité de faire une erreur de type II diminue.

Le troisième paramètre est la taille de l'effet à mettre en évidence. La taille de l'effet est définie comme la différence minimale d'effet que l'investigateur souhaite détecter. Elle est également appelée la différence d'effet cliniquement pertinente.

Enfin, le dernier paramètre de base est la variabilité. Cette notion de variabilité est liée à la variance de l'estimateur de l'effet traitement et diffère en fonction du type d'études. En effet, pour les essais *N of 1*, la variabilité qui est prise en compte est la variabilité intra-patient. Tandis que dans le cas des essais *stepped wedge*, le terme de variabilité considéré est la variabilité inter-groupe, entre autres. Ces notions seront développées en détails par la suite. Il est cependant important de comprendre qu'en général, lorsque cette variabilité augmente, la taille d'échantillon requise augmente également. Il est simple d'imaginer, par exemple, une étude RCT de l'effet de deux traitements sur la tension artérielle sur deux groupes d'individus. Si la variance de la variable d'intérêt, la tension artérielle, est importante, cela signifie que la

tension artérielle est fort différente d'un individu à l'autre. Dans ce cas, il sera nécessaire de considérer une grande taille d'échantillon afin que la différence des valeurs des traitements entre les groupes ne soit pas masquée par la différence qui existe naturellement entre les individus. Généralement, les investigateurs utilisent une valeur provenant d'études antérieures.

Les calculs de taille d'échantillon pour chacun des designs vont être présentés. Les mêmes conditions que celles considérées dans l'exposition des modèles sont supposées, c'est-à-dire une variable réponse continue et normalement distribuée.

4.1 Essais croisés

Dans son livre, Senn propose deux approches différentes pour calculer la taille d'échantillon (Senn S. , 2002). La première se base sur un taux de précision donné. Tandis que la deuxième est plutôt basée sur la puissance désirée. Cette dernière est la méthode majoritairement utilisée dans la littérature pour le calcul de la taille d'échantillon des essais croisés. Cette section se concentrera donc sur la méthode basée sur la puissance.

Senn présente le calcul de la taille d'échantillon dans la situation d'un essai croisé AB/BA balancé avec un total de n patients, répartis en deux groupes de $n/2$ par séquence. Les aspects théoriques utilisés sont ceux présentés au point 3.1.3, lors de l'estimation de l'effet traitement en considérant un effet période.

Une méthode connue des essais cliniques pour trouver une taille d'échantillon est l'utilisation d'une puissance donnée pour mettre en évidence une différence de traitements souhaitée. Cette approche consiste à considérer la taille d'échantillon en termes de tests d'hypothèses (Senn S. , 2002). Un test t ayant pour but de tester s'il existe une différence entre deux traitements est considéré. L'hypothèse nulle de ce test t est que la différence entre les traitements, telle que définie par la vraie différence de traitements non connue notée θ, est de 0. L'hypothèse alternative est que cette différence est différente de 0.

$$H_0: \theta = 0$$

$$H_1: \theta \neq 0$$

En considérant $\hat{\theta}$ comme un estimateur de θ et $\hat{\sigma}_\theta$ comme étant l'estimateur de son écart type basé sur dl degrés de liberté et $t_{\alpha/2,dl}$, la valeur critique de la distribution de t avec dl degrés de liberté, tel que $P\left(t \geq t_{\alpha/2,dl}\right) = \alpha/2$, alors l'hypothèse nulle n'est pas rejetée lorsque :

$$-t_{\alpha/2,dl} < \hat{\theta}/\hat{\sigma}_\theta < t_{\alpha/2,dl} \qquad (14)$$

A partir de ce test, les deux erreurs développées dans l'introduction peuvent être commises. La première erreur est l'erreur de type I liée au niveau de significativité α. En imaginant que l'hypothèse alternative soit fausse, si la statistique de test t $\hat{\theta}/\hat{\sigma}_\theta$ tombe hors des limites des valeurs critiques alors ce test mènera au rejet de l'hypothèse nulle. L'hypothèse nulle est alors

rejetée en faveur de l'hypothèse alternative alors qu'en réalité cette dernière est fausse. Cette erreur est l'erreur de type I.

(*14*) peut être réécrite de manière équivalente comme :

$$-\hat{\sigma}_\theta t_{\alpha/2,dl} < \hat{\theta} < \hat{\sigma}_\theta t_{\alpha/2,dl} \quad (15)$$

En supposant cette fois que l'hypothèse nulle est fausse et donc que l'hypothèse alternative est vraie, si l'estimateur de l'effet traitement, $\hat{\theta}$, tombe dans les limites données par (*15)* alors l'hypothèse nulle ne sera pas rejetée. Cette erreur est l'erreur de type II et sa probabilité, sachant que H_1 est vraie est notée β. Pour toutes valeurs de θ non nulles, β peut être calculé comme :

$$P\left(-\hat{\sigma}_\theta t_{\alpha/2,dl} < \hat{\theta} < \hat{\sigma}_\theta t_{\alpha/2,dl}\right) = P\left(-t_{\alpha/2,dl} < \frac{\hat{\theta}/\sigma_\theta}{\hat{\sigma}_\theta/\sigma_\theta} < t_{\alpha/2,dl}\right) \quad (16)$$

Quelques précisions sont à faire sur le terme central de l'égalité de droite. Le numérateur $\hat{\theta}/\sigma_\theta$ a une distribution normale avec une moyenne de δ/σ_θ, avec δ la différence de l'effet traitement ciblée et une variance de 1. Le dénominateur, quant à lui, $\hat{\sigma}_\theta/\sigma_\theta$ est en réalité la racine carrée d'un khi-carré divisée par ses degrés de liberté dl. Étant donné que le numérateur et le dénominateur sont indépendants et que δ est égal à 0 sous H_0, le terme central de (*16)* a une distribution *t* de student avec dl degrés de liberté. Toutefois, sous H_1, δ n'est plus égal à 0 et le terme central a une distribution *t non centrale* avec dl degrés de liberté. Le paramètre de non centralité est $\gamma = \delta/\sigma_\theta$ (Senn S. , 2002). En considérant la variable aléatoire, désignée comme $t_{dl}(\gamma)$, alors (*16)* peut se réécrire comme :

$$P\left(-t_{\alpha/2,dl} < t_{dl}(\gamma) < t_{\alpha/2,dl}\right) \quad (17)$$

En pratique pour calculer β, les valeurs considérées de γ sont uniquement les valeurs positives telles que :

$$P\left(t_{dl}(\gamma) < t_{\alpha/2,dl}\right) \quad (18)$$

Senn développe deux arguments afin de justifier ce choix (Senn S. , 2002). Premièrement, la différence entre (*17)* et *(18)* est faible. Bien que le β de (*18)* sera plus élevé, il est également plus simple à obtenir. Étant donné que, sous H_1, $t_{dl}(\gamma)$ suit une distribution t non centrée, cela implique que la probabilité d'être inférieure à $-t_{\alpha/2,dl}$ est très faible. Cette probabilité se rapproche de plus en plus de 0 au fur et à mesure que γ augmente. De ce fait, la différence entre (*17*) et (*18*) est considérée comme minime. Deuxièmement, la seule différence entre les deux équations est que (*17)* permet de détecter une différence de traitement spécialement dans le cas où le traitement considéré comme le moins bon serait évalué comme plus efficace. En termes d'hypothèses, l'erreur de type II n'est pas commise. Tandis que dans la pratique, cela pourrait amener à une erreur grave. C'est pour ces raisons que les valeurs considérées de γ sont uniquement les valeurs positives (*18*).

Comme développé lors de l'introduction, il est possible de parler en termes de puissance et non plus d'erreur de type II. Pour rappel, la puissance d'un test est la probabilité de mettre en

évidence une différence d'effet lorsque l'hypothèse alternative est vraie et correspond à $1 - \beta$. L'expression en termes de puissance est alors :

$$1 - \beta = P\left(t_{dl}(\gamma) > t_{\alpha/2,dl}\right) \quad (19)$$

A partir de ce calcul de puissance, il est possible d'obtenir un calcul de la taille d'échantillon. Pour se faire, les scientifiques doivent définir une valeur pour α et une valeur pour la puissance du test, $1 - \beta$. Ils doivent également établir une valeur de différence de traitement ciblée, δ. Il est également nécessaire de définir une valeur pour la variance de la 'cross over difference', ${\sigma_{(2)}}^2$. Cette valeur est généralement choisie à partir des résultats de précédentes études.

Une fois que tous ces paramètres sont définis, il reste encore une difficulté. En effet, dans (*19)*, plusieurs paramètres dépendent directement de *n*. Ces derniers sont $dl = n - 2$, ${\sigma_\theta}^2 = {\sigma_{(2)}}^2/n$ et $\gamma = n^{1/2}\delta/\sigma_{(2)}$. Cependant, une approximation de (*19)* peut être obtenue à l'aide d'une distribution normale. L'expression devient :

$$1 - \beta \cong P\left(Z(n^{1/2}\delta/\sigma_{(2)}, 1) > z_{\alpha/2}\right) \quad (20)$$

Avec $Z(a, b)$ une variable aléatoire normale de moyenne a et de variance b, et $z_{\alpha/2}$, le percentile $\alpha/2$ d'une distribution normale $N(0,1)$ tel que $P\left(Z(0,1) > z_{\alpha/2}\right) = \alpha/2$.

(*20)* peut alors se réécrire comme :

$$1 - \beta \cong P\left(Z(0,1) > z_{\alpha/2} - n^{1/2}\delta/\sigma_{(2)}\right)$$

Tel que :

$$-z_\beta \cong z_{\alpha/2} - n^{1/2}\delta/\sigma_{(2)}$$

Ce qui mène au calcul de la taille d'échantillon suivant (Senn S. , 2002):

$$n \cong \left(z_{\alpha/2} + z_\beta\right)^2\left(\sigma_{(2)}/\delta\right)^2 \quad (21)$$

Senn utilise un exemple pour illustrer le calcul d'échantillon (Senn S. , 2002). Ce dernier est le suivant :

Des chercheurs désirent mettre en place un essai croisé AB/BA pour comparer deux bronchodilatateurs différents. La différence de traitement ciblée est de $30\ l/min$ de débit expiratoire maximal avec une puissance de 80 % et un α de 5 %. L'écart type de la 'cross over difference' est de $45\ l/min$.

Les paramètres à disposition sont donc $\alpha = 0.05$, $\beta = 0.2$, $\delta = 30\ l/min$ et $\sigma_{(2)} = 45\ l/min$. Les seuls paramètres manquants à (*21)* sont les $z_{\alpha/2}$ et z_β. Ces deniers sont obtenus grâce à la table de la distribution normale et correspondent à $z_{\alpha/2} = 1.96$ et $z_\beta = 0.84$. L'expression de (*21)* devient alors :

$$n \cong (1.96 + 0.84)^2(45/30)^2 \cong 18$$

Une taille d'échantillon de 18 individus est donc obtenue pour cette étude.

Afin de vérifier si cette taille d'échantillon permet bien de répondre à la demande en termes de puissance, il est possible de recalculer la puissance à partir de la taille d'échantillon obtenue. Pour se faire, l'expression utilisée est (*19)*. Le degré de liberté correspond à $df = n - 2 = 18 - 2 = 16$. Sur base des tables de la distribution t, $t_{0.025,16}$ est égal à 2.120. Le paramètre de non centralité γ est égal à $n^{1/2}\delta/\sigma_{(2)} = 18^{1/2}(30/45) = 2.83$. La puissance du test est donc de $P(t_{0.025,16,2.83} > 2.120) = 0.76$. Étant donné que la puissance attendue était de 0.80, il est nécessaire d'augmenter la taille d'échantillon. En essayant une taille d'échantillon de 20 individus, la puissance est alors de 0.80. Une taille d'échantillon de 20 individus est donc nécessaire pour répondre aux conditions désirées pour cette étude.

4.2 Essais *stepped wedge*

L'utilisation des essais *stepped wedge* est en augmentation depuis quelques années. Cependant, les scientifiques ne sont pas toujours d'accord sur la façon de calculer la taille d'échantillon. Étant donné que les essais *stepped wedge* sont des essais qui sont menés sur plusieurs périodes et dans lesquels les groupes passent progressivement de la situation de contrôle à la situation de traitement tout au long de l'étude, le temps est un facteur de confusion particulièrement important. Ce dernier devrait donc être inclus dans le calcul de la taille d'échantillon pour les essais *stepped wedge*. Des chercheurs ont publié une revue systématique en 2016 afin d'évaluer les méthodes utilisées par les scientifiques pour calculer la taille d'échantillon de 60 articles utilisant le design *stepped wedge* (Martin, Taljaard, Girling, & Hemming, 2016). Les deux méthodes les plus utilisées par les chercheurs qui permettent de tenir compte des effets temporels étaient la méthode Hussey et Hughes et la méthode de Woertman. Étant donné que ces deux méthodes sont les principales utilisées, elles sont développées en détails ci-dessous.

4.2.1 Méthode de Hussey et Hughes

Hussey et Hughes proposent d'évaluer la taille d'échantillon de l'essai en termes de puissance (Hussey & Hughes, 2007). En effet, ils ne fournissent pas un calcul de la taille d'échantillon mais bien un calcul de la puissance à partir des conditions de l'étude (nombre de groupes, nombre de périodes, ...). Cette méthode peut être utilisée dans le cas des essais *stepped wedge cross-sectional*. La méthode de Hussey et Hughes est à l'heure actuelle la méthode la plus utilisée pour tenir compte de l'effet du temps.

A partir de (*9*) présenté au point 3.3.1.2, un des objectifs serait de réaliser un test de Wald ayant comme hypothèse nulle H_0: $\alpha_{1,j} = 0$ et comme hypothèse alternative H_1: $\alpha_{1,j} = \delta$ en utilisant un essai *stepped wedge* constitué de *G* groupes et de K périodes. Pour rappel, le paramètre $\alpha_{1,j}$ correspond à l'effet traitement et δ est l'effet traitement ciblé.

Une approximation de la puissance d'un tel test bilatéral de taille α est (Hussey & Hughes, 2007):

$$1 - \beta \cong \phi \left(\frac{\delta}{\sqrt{Var(\hat{\alpha}_{1,j})}} - Z_{1-\alpha/2} \right)$$

Dans cette expression, ϕ est la fonction de distribution cumulative d'une normale centrée réduite.

Lors de l'exposition des différents modèles du chapitre précédent, les méthodes d'estimation des paramètres ainsi que de leurs variances dans le cas des modèles mixtes n'ont pas été présentées. Ces méthodes étant complexes, les auteurs proposent une formule plus simple dans le but d'estimer la variance de l'estimateur de l'effet traitement, $Var(\hat{\alpha}_{1,j})$. Cette expression analytique s'applique pour des modèles sous la forme de (9) dans les essais *stepped wedge cross-sectional* et en considérant la dummy variable $treat_j$ prenant la valeur de 1 si le groupe est sous les conditions de traitement et 0 sinon (Hussey & Hughes, 2007). En supposant un nombre égal d'individus par groupe et par période avec un nombre de groupes, noté G et un nombre de périodes, noté K alors :

$$Var(\hat{\alpha}_{1,j}) = \frac{G{\sigma_{wg}}^2({\sigma_{wg}}^2 + K{\sigma_g}^2)}{(GU - W){\sigma_{wg}}^2 + (U^2 + GKU - KW - GV){\sigma_g}^2}$$

Dans cette expression, les auteurs définissent U, W et V comme (Hussey & Hughes, 2007) :

$$U = \sum_{g=1}^{G} \left(\sum_{k=1}^{K} treat_j \right) ; \; W = \sum_{k=1}^{K} \left(\sum_{g=1}^{G} treat_j \right)^2 ; \; V = \sum_{g=1}^{G} \left(\sum_{k=1}^{K} treat_j \right)^2$$

Dans ce contexte, U correspond au nombre d'individus par groupe qui ont été soumis à la situation de traitement tout au long des différentes périodes. W représente le carré de la somme des individus des groupes soumis au traitement par période. Les sommes sur toutes les périodes sont ensuite réalisées. V représente, quant-à-lui, le carré de la somme de tous les individus soumis au traitement dans un groupe pour chacune des périodes. Les sommes sur tous les groupes sont ensuite réalisées. Pour rappel, ${\sigma_g}^2$ représente la variance inter-groupe. Tandis que ${\sigma_{wg}}^2$ correspond à σ^2/n_g avec σ^2, la variance intra-groupe et n_g, le nombre d'individus par groupe par période.

Un paramètre important dans les essais *stepped wedge* est le choix du nombre de périodes. Les auteurs ont montré qu'au plus le nombre de périodes augmente, au plus la puissance va augmenter (Hussey & Hughes, 2007).

Le principe de cette méthode est donc d'estimer une puissance par simulation en faisant varier différentes caractéristiques de l'étude désirée. Il est donc nécessaire de fournir des valeurs de base pour les deux variances, intra- et inter- groupe ainsi qu'une valeur moyenne de base de la variable réponse. Ces informations peuvent être obtenues à partir d'études

antérieures. Des scientifiques ont réalisé une étude dont l'objectif était d'améliorer le traitement des escarres grâce à un renforcement des soins multidisciplinaires au Canada (Stern, et al., 2014). Ces derniers ont calculé la taille d'échantillon nécessaire pour réaliser leur étude à l'aide de la méthode d'Hussey et Hughes en faisant varier l'effet traitement ciblé et le nombre d'établissements. Ils ont simulé différentes données et ont estimé la puissance comme étant la proportion d'effet traitement significatif. Une puissance de 80% était considérée comme adéquate.

4.2.2 Méthode de Woertman

Certains auteurs ont proposé une deuxième méthode pour calculer la taille d'échantillon dans les essais *stepped wedge*. Ces derniers proposent une formule obtenue sur base du calcul de la taille d'échantillon des essais en groupes parallèles (Woertman, et al., 2013). Bien que cette deuxième méthode soit moins utilisée comparée à celle de Hussey et Hughes, elle représente un intérêt particulier puisqu'elle permet de tenir compte de l'effet du temps dans le calcul de la taille d'échantillon.

Les auteurs considèrent un essai *stepped wedge* dans lequel le nombre de groupes qui passent de la situation de contrôle à la situation de traitement à chaque étape reste identique et le nombre de périodes pendant lesquelles les mesures sont récoltées après chaque passage du contrôle au traitement reste constant.

Avant de présenter cette deuxième méthode, le calcul de la taille d'échantillon des essais en groupes parallèles va être présenté puisque le calcul proposé par Woertman est basé sur ce dernier. Plusieurs auteurs ont développé différentes formules afin de calculer une taille d'échantillon dans les essais de type RCT en groupes parallèles avec une randomisation individuelle. Une des formules souvent utilisée, pour une mesure continue, est la suivante (Hemming, Girling, Sitch, Marsh, & Lilford, 2011) :

$$n_{RCT} = 2\sigma^2 \left\{ \frac{\left(z_{\alpha/2} + z_{\beta}\right)^2}{\delta^2} \right\}$$

n_{RCT} correspond au nombre d'individus à inclure dans chacun des bras de l'étude. σ^2 correspond à la variance de la mesure de la variable d'intérêt. Cette variance est considérée comme étant identique dans chacun des bras, de contrôle et de traitement, de l'étude. Le paramètre δ correspond, comme précédemment, à la différence d'effet ciblée.

Dans le cas des essais cliniques contrôlés randomisés en groupes (*Cluster Randomised Controlled Trials* ou CRCT en anglais), la randomisation n'est plus individuelle mais bien réalisée par groupe. Le calcul de la taille d'échantillon de base n_{RCT} est alors ajusté et multiplié par un facteur de correction, appelé en anglais le *'design effect'*. Ce dernier correspond à $\left[1 + \left(n_g - 1\right)\rho\right]$, avec n_g correspondant au nombre de sujets par groupe et ρ, au coefficient de corrélation intra-groupe (Hemming, Girling, Sitch, Marsh, & Lilford, 2011). Ce dernier correspond au rapport entre la variabilité inter-groupe et la somme des variabilités intra- et inter-groupe. Ce coefficient est défini comme la proportion de variabilité de la variable

réponse qui peut être expliquée par la variance inter-groupe (Campbell, Fayers, & Grimshaw, 2005).

Woertman propose d'utiliser une approche similaire afin de corriger le calcul d'échantillon de base pour l'appliquer dans le cas précis des essais *stepped wedge* (Woertman, et al., 2013).

Le facteur de correction (fc) pour les essais *stepped wedge* proposé par Woertman est :

$$fc = \frac{1 + \rho\left(SDn_g + Bn_g - 1\right)}{1 + \rho\left(\frac{1}{2}SDn_g + Bn_g - 1\right)} \times \frac{3(1-\rho)}{2D\left(S - \frac{1}{S}\right)}$$

Dans cette expression, le paramètre S représente la séquence donc le nombre de passages des groupes de l'état de contrôle à l'état de traitement. Dans l'exemple de la *figure 8*, 4 séquences étaient représentées. Le paramètre B représente le nombre de périodes au cours desquelles les mesures de référence sont récoltées. Pour rappel, dans un essai *stepped wedge*, tous les groupes sont dans une situation de contrôle en début d'étude donc ces mesures représentent des mesures de référence. Dans la présentation des analyses des essais *stepped wedge (figure 8)*, le paramètre B était considéré comme étant égal à 1 puisque les mesures de référence chez tous les groupes étaient récoltées au cours de la première période seulement. Enfin, le paramètre D représente le nombre de périodes au cours desquelles les mesures sont récoltées après chaque passage d'une situation à une autre. Dans la présentation des analyses des essais *stepped wedge (figure 8)*, le paramètre D était égal à 1 puisqu'une seule période de récolte des mesures était considérée après chacun des passages de la situation de contrôle à la situation de traitement des différents groupes.

A présent, le calcul de la taille d'échantillon d'un essai *stepped wedge* peut être obtenu en multipliant la formule non ajustée par le facteur de correction : $n_{sw} = n_{RCT} \times fc$. Il est important de remarquer que n_{RCT} correspond généralement à la taille d'un bras d'une étude RCT en groupes parallèles alors que le calcul de n_{sw} rend un nombre total de sujets nécessaire à l'étude. Il ne faut donc pas oublier de multiplier n_{RCT} par le nombre de bras de l'étude avant de l'appliquer dans la formule. Ce nombre correspond donc à 2 dans un essai *stepped wedge* mis en place pour étudier la différence entre un contrôle et un traitement.

A partir du nombre total d'individus nécessaire à l'étude n_{sw}, le nombre optimal G de groupes correspond à n_{sw}/n_g, où n_g correspond au nombre d'individus par groupe. De plus, le nombre de groupes passant de l'état de contrôle à l'état de traitement à chaque période s'obtient en divisant le nombre de groupes G par le nombre de périodes K.

Étant donné que le facteur de correction est fortement influencé par le choix de plusieurs paramètres, il est primordial de connaître les différents degrés d'influence de ces derniers (Woertman, et al., 2013). Premièrement, augmenter le nombre de périodes de récolte après chaque passage du contrôle au traitement, D, diminue le facteur de correction. La même constatation est réalisée lorsque le nombre de passages du contrôle au traitement, S, ou que le nombre de périodes de récolte des mesures de référence, B, augmente. Ainsi

l'augmentation d'un des paramètres susmentionnés engendre une diminution de la taille d'échantillon requise.

Le facteur de correction dépend également du coefficient de corrélation intra-groupe. En effet, lorsque ce coefficient augmente, le facteur de correction a tendance à diminuer. Bien que les auteurs mettent en évidence une légère augmentation du facteur de correction pour des valeurs de coefficient inférieures à 0.05. Ce paramètre dépend des caractéristiques des groupes d'individus étudiés et son estimation peut être appuyée, comme évoqué précédemment, par des études antérieures.

4.3 Essais *N of 1*

Jusqu'à présent, les essais *N of 1* réalisés sur des patients ont principalement comme objectif d'étudier l'efficacité des traitements sur certains patients précis. Cependant, certains auteurs envisagent la possibilité d'utiliser ce type d'essai sur la population (Zucker, et al., 1997).

A l'heure actuelle, peu d'auteurs se sont intéressés au calcul de la taille d'échantillon pour les essais *N of 1*. Bien que récemment l'auteur Senn a imaginé un calcul de la taille d'échantillon dans le cas spécial des essais *N of 1* (Senn S. , 2019). Senn ne fournit pas un calcul d'échantillon propre à ce type d'essais mais plutôt un dérivé de celui vu pour les essais croisés (Senn S. , 2019). L'auteur considère un essai *N of 1* dans lequel les traitements administrés à chacun des patients sont randomisés par bloc pour un nombre de blocs $b > 2$. L'auteur émet également les hypothèses que la maladie étudiée est relativement stable et qu'un potentiel effet *carry over* est contrôlé par une période de *wash-out*.

L'auteur propose deux méthodes différentes liées à deux modèles. La première méthode est proposée sur base d'un modèle mixte dans lequel un effet traitement spécifique à chaque patient est considéré comme aléatoire. Tandis que la deuxième méthode est basée sur un modèle dans lequel tous les paramètres sont considérés comme fixes. L'idée est de définir la variance de l'estimateur de l'effet traitement et de l'insérer dans l'expression développée pour calculer la taille d'échantillon des essais croisés (21). Les deux méthodes sont détaillées ci-dessous.

Tout d'abord, l'auteur se base sur les analyses présentées au point 3.4.1.1 pour définir la variance de l'estimateur.

Pour rappel, à partir de (*13*), la variance de la moyenne des différences d'un patient $\bar{y}_{i.}$ était (Araujo, Julious, & Senn, 2016):

$$Var(\bar{y}_{i.}) = {\sigma_x}^2 + ({2\sigma_w}^2/B)$$

La variance de la moyenne des différences de tous les patients donc de l'estimateur de l'effet traitement de tous les patients était :

$$Var(\bar{Y}) = \frac{{\sigma_x}^2 + \ ({2\sigma_w}^2/B)}{n}$$

A partir de cet estimateur, un test t peut être réalisé. L'hypothèse nulle de ce test est que l'effet traitement est égal à 0, donc qu'il n'y pas de différence entre les deux traitements. Son hypothèse alternative est que l'effet traitement est différent de 0. Les degrés de liberté du test *t* pour l'estimation de l'effet traitement sur l'ensemble des patients sont alors de $n-1$.

L'auteur envisage un calcul de la taille d'échantillon à l'aide de la formule développée pour les essais croisés (*21*). La variance de la moyenne des différences ${\sigma_x}^2 + (2{\sigma_w}^2/B)$ est implémentée dans l'expression (21). L'expression (21) devient alors :

$$n \cong \frac{\left(z_{\alpha/2}+z_{\beta}\right)^2({\sigma_x}^2+(2{\sigma_w}^2/B))}{\delta^2} \qquad (22)$$

De même que pour les essais croisés, la puissance peut être vérifiée une fois que la taille d'échantillon est calculée. Le calcul de la puissance peut être obtenu à l'aide d'une distribution *t non centrale* avec $(n-1)$ degrés de liberté. Le paramètre de non centralité est :

$$\gamma = \frac{\delta}{\sqrt{\frac{{\sigma_x}^2 + (2{\sigma_w}^2/B)}{n}}}$$

avec δ, la différence d'effet ciblée.

Senn propose un exemple fictif simple afin d'illustrer sa méthode (Senn S. , 2019). Un essai *N of 1* est mis en place dans le but de comparer deux traitements administrés de manière répétée à chaque patient. L'administration des traitements est répétée sur trois blocs. La différence ciblée entre les traitements est de 1 avec une puissance de 80 % et un α de 5 %. La variance de la différence des traitements entre les patients est considérée comme étant de 1. La variance intra-patient est, quant-à-elle, égale à 4.

Les paramètres à disposition sont donc $\alpha = 0.05$, $1-\beta = 0.8$,$B = 3$, $\delta = 1$, ${\sigma_w}^2 = 4$ et ${\sigma_x}^2 = 1$. Le calcul de la taille d'échantillon des essais *N of 1* est réalisé à partir de (*22).* La variance de l'estimateur est :

$${\sigma_x}^2 + \left(2{\sigma_w}^2/B\right) = 1 + (2 * 4/3) = 3.67$$

Le calcul de la taille d'échantillon est donc :

$$n \cong \frac{(1.96 + 0.84)^2(3.67)}{1^2} \cong 29$$

Une taille d'échantillon de 29 individus est obtenue pour cette étude.

Les degrés de liberté correspondent à $dl = n - 1 = 29 - 1 = 28$. Sur base des tables de la distribution *t*, $t_{0.025,28}$ est égal à 2.048. Le paramètre de non centralité γ est égal à :

$$\gamma = \frac{1}{\sqrt{3.67/29}} = 2.81$$

La puissance du test est donc de $P\left(t_{0.025,28,2.81} > 2.048\right) = 0.77$. Étant donné que la puissance attendue était de 0.80, il est nécessaire d'augmenter la taille d'échantillon. En essayant une

taille d'échantillon de 31 individus, la puissance est alors de 0.80. Une taille d'échantillon de 31 individus est donc nécessaire pour répondre aux conditions désirées pour cette étude.

L'auteur propose une deuxième méthode. Cette dernière est à utiliser lorsque les effets des traitements sont étudiés spécifiquement sur chacun des patients de l'étude ou lorsque ${\sigma_x}^2$ est considéré comme étant nul. Cette dernière hypothèse est observée lorsque l'effet traitement est le même pour tous les patients.

Pour cette deuxième méthode, l'auteur précise que le test de l'effet traitement n'a plus un degré de liberté $n-1$ mais bien de $n(B-1)$. En effet, dans ce deuxième cas, une variance pour la moyenne des différences sera estimée pour chacun des patients, séparément. Chacune de ces estimations aura donc $B-1$ degrés de liberté. Une moyenne générale de toutes les moyennes de chacun des patients est ensuite estimée. Étant donné qu'il existe n patients, la variance de cette dernière est basée sur un total de $n(B-1)$ degrés de liberté (Senn S. , 2019).

Étant donné que ${\sigma_x}^2 = 0$, la variance de la moyenne des différences de tous les patients se résume à (Senn S. , 2019) :

$$\frac{2{\sigma_w}^2}{nB}$$

L'auteur propose alors un calcul de la taille d'échantillon à partir de (*21*) en utilisant cette fois cette nouvelle estimation de la variance (Senn S. , 2019). Le calcul de taille d'échantillon est comme suit (Senn S. , 2019) :

$$n \cong \frac{\left(z_{\alpha/2}+z_{\beta}\right)^2 2{\sigma_w}^2}{B\,\delta^2} \tag{23}$$

Comme précédemment, un calcul de puissance peut être réalisé afin de vérifier que la taille d'échantillon correspond bien aux exigences désirées. Le calcul de la puissance peut être obtenu à l'aide d'une distribution *t non centrale* avec $n(B-1)$ degrés de liberté. Le paramètre de non centralité est :

$$\gamma = \frac{\delta}{\sqrt{\frac{2{\sigma_w}^2}{Bn}}}$$

Cette deuxième méthode peut être illustrée en reprenant l'exemple utilisé pour présenter la première méthode. Pour rappel, les paramètres à disposition étaient $\alpha = 0.05$, $1-\beta = 0.8$, $B = 3$, $\delta = 1$, ${\sigma_w}^2 = 4$ et ${\sigma_x}^2 = 1$. En supposant cette fois, une variabilité de l'effet traitement entre les patients comme étant nulle, ${\sigma_x}^2 = 0$. La variance de l'estimateur devient :

$$\frac{2{\sigma_w}^2}{B} = \frac{2 \times 4}{3} = 2.67$$

A partir de (*23*), le calcul de la taille d'échantillon est donc :

$$n \cong \frac{(1.96 + 0.84)^2\ 2.67}{1^2} \cong 21$$

Le degré de liberté correspond à $dl = n(B-1) = 21(3-1) = 42$. Sur base des tables de la distribution *t*, $t_{0.025,42}$ est égal à 2.021. Le paramètre de non centralité γ est égal :

$$\gamma = \frac{1}{\sqrt{2.67/21}} = 2.80$$

La puissance du test est donc de $P\left(t_{0.025,42,2.80} > 2.021\right) = 0.78$. Étant donné que la puissance attendue était de 0.80, il est nécessaire d'augmenter la taille d'échantillon. En essayant une taille d'échantillon de 22 individus, la puissance est alors de 0.80.

5 Discussion

Plusieurs modèles mixtes différents ont été présentés afin d'analyser les essais croisés, les essais *stepped wedge* et les essais *N of 1*. Les modèles mixtes, à covariance pattern ou à effets aléatoires, sont des modèles qui offrent la possibilité de tenir compte de la corrélation entre les données, causée par les designs eux-mêmes. Le choix du type de modèle à utiliser va dépendre des objectifs de l'étude et des possibilités. Ce choix dépendra également du nombre de paramètres estimables en fonction du nombre d'observations à disposition. Certains des modèles à covariance pattern peuvent avoir un nombre de paramètres à estimer relativement important. Les modèles à effets aléatoires contiennent, quant-à-eux, un nombre de paramètres moins important. De plus, ce nombre restera identique peu importe le nombre d'individus présents dans l'étude. Le choix du type de modèle et de la structure de corrélation dépendra également de la façon dont les données sont récoltées. En effet, certaines structures sont à utiliser lorsque les données sont espacées de manière régulière. Quelques auteurs ont développé des modèles à covariance pattern dans lesquels des effets aléatoires sont ajoutés. Ces derniers ont été présentés lors de l'analyse des essais *stepped wedge closed cohort*.

Les méthodes d'analyse ainsi que le calcul de la taille d'échantillon des essais croisés sont connus et utilisés par de nombreux scientifiques depuis plusieurs années. L'intérêt pour les essais *stepped wedge* et les essais *N of 1* est cependant beaucoup plus récent. Les méthodes et les techniques d'analyse développées sont donc beaucoup moins nombreuses, principalement pour les essais *N of 1*.

De nombreux auteurs se sont déjà intéressés à l'utilisation et l'analyse statistique des essais *stepped wedge*. Différentes méthodes d'analyse des essais *stepped wedge cross-sectional* et *closed cohort* ont été présentées. Ce travail a montré que les essais *stepped wedge closed cohort* sont plus complexes à analyser, comparés aux essais *cross-sectional*. En effet, le nombre de paramètres dans les modèles des essais *closed cohort* est souvent plus élevé. De plus, les structures de corrélation ont tendance à être plus complexes. Cependant, les essais *stepped wedge closed cohort* offrent l'avantage d'étudier la variabilité des mesures prises chez un même patient et ainsi d'obtenir des mesures plus précises. Les essais *closed cohort* sont plus utilisés que les essais *stepped wedge cross-sectional*. Il existe un troisième type d'essais *stepped wedge*, les *open cohort*. Dans ces derniers, des patients peuvent se rajouter à la cohorte à chaque période. Ce type d'essais *stepped wedge* implique une structure de corrélation des données et des analyses plus complexes. Ces essais n'ont pas été développés dans ce travail.

Il est clair que les articles de la littérature se concentrent surtout sur l'effet de groupe ainsi que l'effet période (Li, Hughes, Hemming, & Taljaard, 2021). En effet, ces deux effets représentent une source de variabilité importante dans les essais *stepped wedge*. Étant donné que ces essais sont réalisés sur des groupes d'individus, il est possible que les mesures récoltées chez les patients provenant d'un même groupe soient corrélées entre elles. Bien

que cet effet n'ait pas été considéré dans le cas des essais croisés, il est tout à fait possible d'imaginer une situation dans laquelle un effet groupe est présent. Par exemple, en imaginant un essai croisé réalisé dans une école sur des élèves provenant de différentes classes, les élèves provenant de la même classe ont un risque d'avoir des données corrélées. Cet effet groupe pourrait également être présent dans les essais en groupes parallèles et dans les études RCT en général. En réalité, lorsqu'un ensemble d'individus provient d'un même groupe (les classes d'une école, des hôpitaux ou des centres de soins …), un tel effet peut être présent. Dans ce cas, il est nécessaire de tenir compte dans le modèle d'un effet groupe. Le fait que les essais *stepped wedge* soient réalisés sur des groupes d'individus représente un énorme avantage. Tous les individus ont le traitement en même temps, il n'y a donc pas de 'contamination' entre les individus et aucun biais de sélection n'est présent. Le deuxième effet, mis en avant dans l'analyse des essais *stepped wedge*, est l'effet période. Comme les essais croisés, les essais *stepped wedge* sont des études dans lesquelles les patients sont suivis au cours de nombreuses périodes. Il est important de considérer l'impact d'un potentiel effet période sur les individus.

Un point important qui a été présenté dans l'analyse des *stepped wedge closed cohort* et qui est peu présent dans la littérature est la modélisation de l'effet traitement (Li, Hughes, Hemming, & Taljaard, 2021). Comme expliqué dans ce travail, le temps depuis la mise en place du traitement pourrait avoir un impact sur l'effet traitement lui-même. Certains des groupes ont reçu le traitement à des périodes plus précoces que d'autres. Il est donc possible que certaines mesures soient réalisées plusieurs périodes après la mise en place du traitement. Différents modèles ont été présentés dans ce travail afin de tenir compte de l'effet potentiel des périodes sur l'effet traitement. Cependant, la modélisation sous cette forme de l'effet traitement n'est pas celle majoritairement utilisée dans la littérature. En effet, l'effet traitement est le plus souvent considéré comme un terme constant qui ne change pas, et ce, peu importe le nombre de périodes écoulées depuis la mise en place du traitement. Bien que cette modélisation ait été vue dans le cas des essais *stepped wedge closed cohort*, elle pourrait être envisagée dans un essai *stepped wedge cross-sectional* en considérant un suivi de tous les patients au cours des mêmes périodes mais une récolte unique de données pour chaque patient.

Dans la littérature scientifique, les recherches en ce qui concerne le calcul de la taille d'échantillon de essais *stepped wedge* sont beaucoup moins nombreuses. La méthode principalement utilisée jusqu'à présent par les chercheurs est la méthode de Hussey et Hughes (Hussey & Hughes, 2007). Bien que cette méthode tienne compte d'un potentiel effet période, cet effet période est considéré comme identique pour tous les groupes (Woertman, et al., 2013). Un deuxième inconvénient de cette méthode est l'hypothèse faite par les auteurs. Ceux-ci considèrent un nombre égal d'individus par groupe et par période dans leur expression de l'estimation de la variance de l'estimateur de l'effet traitement. Au vu du design des essais *stepped wedge*, récolter un même nombre d'observations par groupe par période peut parfois ne pas être réalisable dans la pratique. Une deuxième méthode, moins utilisée, a

également été présentée, la méthode de Woertman (Woertman, et al., 2013). L'auteur précise que le facteur de correction sera généralement inférieur à 1 ce qui implique qu'un nombre moins élevé d'individus est nécessaire pour réaliser un essai *stepped wedge* comparé à un essai en groupes parallèles. Comme expliqué précédemment, cette réduction de taille d'échantillon requise peut provenir du fait que les groupes sont considérés comme leur propre témoin étant donné que tous passent par la situation de contrôle et de traitement. De plus, les auteurs considèrent, comme pour la méthode de Hussey et Hughes, un nombre d'individus identique pour chaque groupe. Ces deux méthodes sont principalement développées pour être utilisées dans le cas des essais *stepped wedge cross-sectional*. En effet, aucune des deux méthodes ne considère des mesures répétées par individu et donc une corrélation entre ces dernières. Dans la littérature, de nombreux auteurs ne référencient pas la méthode utilisée pour calculer la taille de l'échantillon dans leurs recherches (Martin, Taljaard, Girling, & Hemming, 2016).

Il est clair qu'il faut garder à l'esprit que les essais *stepped wedge* sont des essais qui sont principalement utilisés lorsqu'il existe des contraintes. En effet, les études *stepped wedge* sont généralement mises en place pour des raisons éthiques, lorsque le traitement doit être administré à l'ensemble des groupes (centres, hôpitaux, écoles…). Elles permettent également d'éviter des contraintes, notamment des contraintes logistiques, par exemple lorsqu'il n'est pas possible d'administrer le traitement à tous les groupes en même temps. L'utilisation de ce type d'études est donc justifiée dans des cas bien particuliers.

En ce qui concerne les essais *N of 1*, peu d'auteurs se sont intéressés à leurs analyses ainsi qu'au calcul de la taille d'échantillon. Certains auteurs ont proposé des modèles statistiques. Parmi ces modèles, nombreux sont ceux qui contiennent un nombre de paramètres très élevé comparé au nombre d'observations contenues dans les essais *N of 1*. En effet, les essais *N of 1* ont été créés dans le but d'étudier l'effet traitement pour un patient spécifique. Dans ce premier contexte, le nombre d'observations disponibles est très réduit. La mise en place d'un modèle statistique pour analyser les données d'un seul patient semble donc compliquée, étant donné qu'il faut un nombre d'observations suffisant pour estimer tous les paramètres du modèle. Le nombre d'observations récoltées pour le patient étudié augmente lorsque le nombre de blocs au cours desquels le patient est suivi est plus important. Cependant, au vu des risques liés à l'administration successive des traitements et à l'inconfort que cette dernière peut engendrer, les essais *N of 1* sont généralement constitués d'un nombre restreint de blocs. Par conséquent, les modèles statistiques ne sont généralement pas mis en place dans les analyses des essais *N of 1* réalisés sur un seul patient.

Dans le cas où l'essai contiendrait plusieurs patients, certains auteurs proposent des modèles statistiques à covariance pattern. Au vu du nombre de paramètres parfois élevé de ce type de modèle, il faut garder à l'esprit qu'ils ne pourront être mis en place que si l'essai *N of 1* contient suffisamment d'individus. Par ailleurs, des auteurs proposent d'utiliser des structures qui n'impliquent pas une corrélation entre les données d'un même patient. La covariance entre les observations d'un même patient est donc considérée comme étant égale à 0. Cependant,

au vu de la structure des essais *N of 1*, le fait de considérer les données d'un même patient comme non corrélées ne semble pas être approprié.

L'administration répétée de plusieurs traitements à un même individu permet d'obtenir une estimation plus précise de l'effet traitement par individu (Zucker, Ruthazer, & Schmid, 2010). Certains auteurs proposent d'utiliser un effet traitement aléatoire dans l'analyse des essais *N of 1* réalisés sur plusieurs individus. La modélisation de l'effet traitement comme un effet aléatoire permet d'obtenir une estimation de la variance de l'effet traitement entre les individus, donc de savoir si cet effet est le même pour tous les individus. Il serait également possible de considérer un effet aléatoire traitement lorsque l'essai est réalisé sur un seul patient. Considérer un tel effet permettrait d'obtenir une estimation de la variabilité de l'effet traitement au cours des différents blocs, pendant lesquels le patient étudié reçoit les différents traitements successivement. Cependant dans la littérature, les auteurs ne considèrent pas des modèles avec un tel effet lorsqu'un seul patient est présent dans l'étude.

Étant donné que les essais *N of 1* peuvent être considérés en réalité comme de multiples essais croisés, ils sont également soumis au risque d'un potentiel effet *carry over*. Certains auteurs envisageaient d'utiliser les essais *N of 1*, dans lesquels l'administration de différents traitements est répétée de nombreuses fois, pour étudier au mieux la présence ou non d'un effet *carry over* (Senn S. , 2002). Cependant, il semblerait qu'un très faible nombre de scientifiques se soient réellement penchés sur la question. Par conséquent, peu de méthodes ont été mises en place pour évaluer l'effet *carry over* dans le cas spécifique des essais *N of 1*. Il est vrai que cet effet est un effet particulièrement difficile à mettre en évidence et à estimer. De nombreux facteurs peuvent influencer son estimation. De plus, les auteurs qui proposent des modèles pour estimer cet effet émettent des hypothèses généralement lourdes comme, par exemple, un effet *carry over* qui ne peut provenir que de la période précédente. Au vu de la structure des essais *N of 1*, ces derniers pourraient donc sous ces hypothèses permettre d'obtenir une meilleure estimation de l'effet *carry over*, comparée à celle obtenue dans le cas d'un essai croisé AB/BA.

Le calcul de la taille d'échantillon proposé dans le cas des essais *N of* 1 présente un gros inconvénient. Dans les calculs de taille d'échantillon en général, une variance de la variable étudiée est nécessaire et cette dernière peut être obtenue grâce à des études préalables disponibles dans la littérature. Dans le calcul d'échantillon des essais *N of 1*, une des variances nécessaire est la variance de l'effet traitement entre les individus. Cependant, cette variance est rarement connue (Senn S. , 2019). En effet, pour obtenir cette estimation, il serait nécessaire de mettre en place des études, comme des essais *N of 1*. Cependant, généralement, de tels essais n'ont pas encore été réalisés.

Les modèles présentés dans ce travail ne contiennent que les effets particulièrement intéressants dans l'analyse de ces différents designs. Cependant, il faut considérer que plusieurs covariables peuvent être ajoutées à chacun des modèles, comme l'âge ou encore le sexe des individus. De plus, les modèles ont été présentés l'un après l'autre afin d'expliquer

précisément chacun des effets et d'éviter de présenter des modèles contenant trop de paramètres. Il est clair que tous les effets présentés pour chaque design dans le cadre de ce travail peuvent être réunis dans un seul modèle et ces modèles peuvent être reconstruits différemment en fonction des objectifs de l'étude.

6 Conclusion

Ce travail a proposé différentes techniques d'analyse et différentes méthodes de calcul de taille d'échantillon pour l'analyse de données des essais croisés, des essais *stepped wedge* ainsi que des essais *N of 1*. Bien que les techniques d'analyse des essais croisés semblent connues et utilisées depuis des années, ce travail a permis de mettre en avant le manque d'articles de la littérature et de recherches concernant les essais *stepped wedge* et *N of 1*.

Alors que les analyses statistiques des essais *stepped wedge* sont développées depuis de nombreuses années, le manque de recherches concernant le calcul de taille d'échantillon se fait ressentir. En effet, il est nécessaire de développer un calcul de la taille d'échantillon qui serait plus simple à mettre en place. De plus, il serait également intéressant d'affiner les recherches afin de développer un calcul de taille d'échantillon pour les essais *stepped wedge closed cohort*.

Il est nécessaire d'augmenter les recherches en ce qui concerne les modèles proposés pour analyser les essais *N of 1*. En effet, les modèles développés et présentés dans la littérature semblent parfois difficiles à mettre en place dans des études *N of 1*. Il serait donc intéressant de développer des modèles plus 'standards' qui peuvent être utilisés en toutes circonstances dans des analyses d'essais *N of 1* contenant plusieurs individus. De plus, il est clair que très peu de recherches ont été réalisées jusqu'à présent en ce qui concerne la taille d'échantillon de ces essais. Il existe donc bien un besoin urgent de développer ces recherches étant donné que les essais *N of 1* sont de plus en plus utilisés. Enfin, il serait intéressant de développer des méthodes afin d'étudier plus précisément l'effet *carry over*, un effet qui semble particulièrement intéressant à étudier dans le cas des essais *N of 1*.

7 Références

Araujo, A., Julious, S., & Senn, S. (2016). Understanding variation in sets of N-of-1 trials. *PloS One*, 11(12), e0167167. https://doi.org/10.1371/journal.pone.0167167.

Bashour, H. N., Kanaan, M., Kharouf, M. H., Abdulsalam, A. A., Tabbaa, M. A., & Cheikha, S. A. (2013). The effect of training doctors in communication skills on women's satisfaction with doctor-woman relationship during labour and delivery: a stepped wedge cluster randomised trial in Damascus. *BMJ Open*, 3(8), e002674. https://doi.org/10.1136/bmjopen-2013-002674.

Brown, C. A., & Lilford, R. J. (2006). The stepped wedge trial design: a systematic review. *BMC Medical Research Methodology*, 6(1), 54. https://doi.org/10.1186/1471-2288-6-54.

Brown, H., & Prescott, R. (2015). *Applied mixed models in medicine (3rd ed.).* John Wiley & Sons.

Campbell, M. K., Fayers, P. M., & Grimshaw, J. M. (2005). Determinants of the intracluster correlation coefficient in cluster randomized trials: the case of implementation research. *Clinical Trials* , 2(2), 99–107. https://doi.org/10.1191/1740774505cn071oa.

Chen, X., & Chen, P. (2014). A comparison of four methods for the analysis of N-of-1 trials. *PloS One*, 9(2), e87752. https://doi.org/10.1371/journal.pone.0087752.

Cushny, A. R., & Peebles , A. R. (1905). The action of optical isomers: II. Hyoscines. *The Journal of Physiology*, 32(5–6), 501–510. https://doi.org/10.1113/jphysiol.1905.sp001097.

Duan , N., Kravitz, R. L., & Schmid, C. H. (2013). Single-patient (n-of-1) trials: a pragmatic clinical decision methodology for patient-centered comparative effectiveness research. *Journal of Clinical Epidemiology*, 66(8 Suppl), S21-8. https://doi.org/10.1. Récupéré sur 66(8 Suppl), S21-8. https://doi.org/10.1

Elbourne, D. R., Altman, D. G., Higgins, J. P., Curtin, F., Worthington, H. V., & Vail, A. (2002). Meta-analyses involving cross-over trials: methodological issues. *International Journal of Epidemiology*, 31(1), 140–149. https://doi.org/10.1093/ije/31.1.140.

Federico, A. C., Heagerty, P. J., Lantos, J., O'Rourke, P., Rahimzadeh, V., Sugarman, J., . . . Magnus, D. (2022, April). Ethical and epistemic issues in the design and conduct of pragmatic stepped-wedge cluster randomized clinical trials. *Contemp Clin Trials*, 115:106703. doi: 10.1016/j.cct.2022.106703. PMID: 35176501. Récupéré sur 115:106703. doi: 10.1016/j.cct.2022.106703. PMID: 35176501.

Fitzmaurice, G., Laird, N., & Ware, J. (2004). Applied Longitudinal Analysis. *John Wiley & Sons*.

Friedman, L. M., Furberg, C. D., DeMets, D. L., Reboussin, D. M., & Granger, C. B. (2015). *Fundamentals of clinical trials (5th ed.).* Fundamentals of clinical trials (5th ed.). Springer International Publishing.

Gupta, K. K., Attri, J. P., Singh, A., Kaur, H., & Kaur, G. (2016). Basic concepts for sample size calculation: Critical step for any clinical trials! *Saudi Journal of Anaesthesia*, 10(3), 328–331. https://doi.org/10.4103/1658-354X.174918.

Guyatt, G., Sackett, D., Taylor, D. W., Ghong, J., Roberts, R., & Pugsley, S. (1986). Determining Optimal Therapy — Randomized Trials in Individual Patients. *New England Journal of Medicine*, 314(14), 889–892. doi:10.1056/nejm198604033141.

Hemming, K. H. (2015). The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. *BMJ (Clinical Research Ed.)*, 350(feb06 1), h391–h391. https://doi.org/10.1136/bmj.

Hemming, K., Girling, A. J., Sitch, A. J., Marsh, J., & Lilford, R. J. (2011). Sample size calculations for cluster randomised controlled trials with a fixed number of clusters. *BMC Medical Research Methodology*, 11(1), 102. https://doi.org/10.1186/1471-2288-11-102.

Hills , M., & Armitage, P. (1979). The two-period cross-over clinical trial. *Br J Clin Pharmacol*, 8(1):7-20. doi: 10.1111/j.1365-2125.1979.tb05903.x. PMID: 552299; PMCID: PMC1429717.

Hussey, M. A., & Hughes, J. P. (2007). Design and analysis of stepped wedge cluster randomized trials. *Contemporary Clinical Trials*, 28(2), 182–191. https://doi.org/10.1016/j.cct.2006.05.007.

Kang, M., Ragan, B. G., & Park, J.-H. (2008). Issues in outcomes research: an overview of randomization techniques for clinical trials. *Journal of Athletic Training*, 43(2), 215–221. https://doi.org/10.4085/1062-6050-43.2.215.

Krone, T., Boessen, R., Bijlsma, S., van Stokkum, R., Clabbers, N., & Pasman, W. (2020). The possibilities of the use of N-of-1 and do-it-yourself trials in nutritional research. *PloS One*, 15(5), e0232680. https://doi.org/10.1371/journal.pone.0232680.

Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2004). *Applied linear statistical models (5th ed.).* McGraw-Hill.

Les éditions Le Robert et Google. (2014, May 6). Récupéré sur Lerobert.com: https://dictionnaire.lerobert.com/google-dictionnaire-fr?param=pharmacodynamique

Li, F., Hughes, J. P., Hemming, K., & Taljaard, M. (2021). Mixed-effects models for the design and analysis of stepped wedge cluster randomized trials: An overview. *Statistical Methods in Medical Research*, 30(2), 612–639. https://doi.org/10.1177/0962280220932962.

Martin, J., Taljaard, M., Girling, A., & Hemming, K. (2016). Systematic review finds major deficiencies in sample size methodology and reporting for stepped-wedge cluster randomised trials. *BMJ Open*, 6(2), e010166. https://doi.org/10.1136/bmjopen-2015-010166.

Molenberghs, G., & Verbeke, G. (2009). Linear Mixed Models for Longitudinal Data. *Springer New York, NY*, https://doi.org/10.1007/978-1-4419-0300-6.

OMS. (2023, October 18). *Maladies non transmissibles. (n.d.).* Récupéré sur Who.int.: https://www.who.int/fr/news-room/fact-sheets/detail/noncommunicable-diseases

Patterson, C. G., Leland, N. E., Mormer, E., & Palmerd, C. V. (2022). Alternative designs for testing speech, language, and hearing interventions: Cluster-randomized trials and stepped-wedge designs. *Journal of Speech, Language, and Hearing Research: JSLHR*, 65(7), 2677–2690. https://doi.org/10.1044/2022_JSLHR-21-00522.

Richard, L. K., & Naihua, D. (2019, August). *Design and Implementation of N-of-1 Trials: A User's Guide*. Récupéré sur Effective Health Care Program, Agency for Healthcare Research and Quality, Rockville, MD: https://effectivehealthcare.ahrq.gov/products/n-1-trials/research-2014-5

Senn, S. (2002). *Cross-over trials in clinical research (2nd ed.).* John Wiley & Sons.

Senn, S. (2019). Sample size considerations for n-of-1 trials. *Statistical Methods in Medical Research*, 28(2), 372–383. https://doi.org/10.1177/0962280217726801.

Senn, S. (2024). The analysis of continuous data from n-of-1 trials using paired cycles: a simple tutorial. *Trials*, 25, 128. https://doi.org/10.1186/s13063-024-07964-7.

Senn, S. J., & Auclair, P. (1990). The graphical representation of clinical trials with particular reference to measurements over time. *Statistics in Medicine*, 9(11), 1287–1302. https://doi.org/10.1002/sim.4780091108.

Stern, A., Mitsakakis, N., Paulden, M., Alibhai, S., Wong, J., Tomlinson, G., . . . Zwarenstein, M. (2014). Pressure ulcer multidisciplinary teams via telemedicine: a pragmatic cluster randomized stepped wedge trial in long term care. *BMC Health Services Research*, 14(1), 83. https://doi.org/10.1186/1472-6963-14-83.

Twisk, J. W. (2021). *Analysis of data from randomized controlled trials: A practical guide (1st ed.).* Springer Nature.

Wasserman, L. (2004). *All of Statistics A Concise Course in Statistical Inference.* Springer.

Woertman, W., de Hoop, E., Moerbeek, M., Zuidema, S. U., Gerritsen, D. L., & Teerenstra, S. (2013). Stepped wedge designs could reduce the required sample size in cluster randomized trials. *Journal of Clinical Epidemiology*, 66(7), 752–758. https://doi.org/10.1016/j.jclinepi.2013.01.009.

Zucker, D. R., Schmid, C. H., McIntosh, M. W., D'Agostino, R. B., Selker, H. P., & Lau, J. (1997). Combining single patient (N-of-1) trials to estimate population treatment effects and to evaluate individual patient responses to treatment. *Journal of Clinical Epidemiology*, 50(4), 401–410. https://doi.org/10.1016/s0895-4356(96)00429-5.

Zucker, D., Ruthazer, R., & Schmid, C. (2010). Individual (N-of-1) trials can be combined to give population comparative treatment effect estimates: methodologic considerations. *Journal of Clinical Epidemiology*, 63(12), 1312–1323. https://doi.org/10.1016/j.jclinepi.2010.04.020.

UNIVERSITÉ CATHOLIQUE DE LOUVAIN
Faculté des sciences
Place des sciences, 2 bte L6.06.01, 1348 Louvain-la-Neuve, Belgique | **www.uclouvain.be/sc**