

Méthodes d'analyse statistique des schémas d'étude en stepped wedge : Application au projet OREANE

Olivier Quintin

▶ **To cite this version:**

Olivier Quintin. Méthodes d'analyse statistique des schémas d'étude en stepped wedge : Application au projet OREANE. Santé publique et épidémiologie. 2021. dumas-03377937

HAL Id: dumas-03377937

https://dumas.ccsd.cnrs.fr/dumas-03377937

Submitted on 14 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Master 2ème année Santé publique - mention Biostatistique
Année universitaire 2020-2021

Mémoire de stage
Du 08/02/2021 au 06/08/2021

Méthodes d'analyse statistique des schémas d'étude en stepped wedge : Application au projet OREANE

Olivier QUINTIN

Maître de stage : Marion KRET (Biostatisticienne)
Equipe : Unité de Soutien Méthodologique à la Recherche clinique et épidémiologique (USMR)
Directeur de l'équipe : Laura Richert
CHU de Bordeaux

RESUME

Le schéma d'étude en stepped wedge fait partie des essais randomisés en clusters. Il présente la particularité d'implémenter une intervention dans tous les groupes ou clusters, de manière décalée à un temps défini par la randomisation. La plupart des essais suivant ce schéma sont regroupés au sein de deux principaux types : transversal et cohorte (ouverte ou fermée). Les différences dans la conception de ces deux schémas impliquent des niveaux de corrélation différents au sein des données qui doivent par conséquent être pris en compte dans l'analyse statistique.

Le but de l'étude OREANE était de déterminer l'impact d'une formation complexe sur le taux d'erreur médicamenteuse des IDE en unité de réanimation néonatale. Différents modèles statistiques ont été explorés dans le but de trouver le modèle le plus adéquat aux données de cette étude suivant un schéma stepped wedge de type cohorte, en accord avec les spécificités de cette dernière. L'utilisation des modèles linéaires généralisés à effets aléatoires prenant en compte les différents niveaux de corrélation (au niveau du cluster et au niveau individuel) a permis de montrer que l'efficacité de l'intervention était dépendante du centre après ajustement sur différentes covariables. Cependant, le faible nombre de clusters disponibles et leur forte hétérogénéité n'a pas permis de conclure à un effet de l'intervention de manière globale.

Mots clés : Stepped wedge, Essais en Clusters Randomisés, Modèles à effets aléatoires, Approches non paramétriques, Analyse intra-période.

ABSTRACT

Stepped wedge trials are a type of Cluster Randomised Trials (CRT). In stepped wedge trials, the clusters are sequentially randomised until the point at which all clusters are exposed to the intervention. Most of these trials are grouped within two types : cross-sectional and cohorts (opened or closed). The differences between these two types imply different levels of correlation which need to be taken into account in the statistical analysis.

The aim of the OREANE study was to determine the effect of a complex training on medication errors in nursing in Neonatal Intensive Care Units. Different statistical methods have been explored in order to find the most adequat model for this cohort stepped wedge study, according to its specificities. The use of Generalized Linear Mixed Model that take into account the different levels of correlation (at the cluster and the individual levels) has shown that the efficacy of the intervention was dependent on the hospital center after adjusting on several covariates. However, the low number of clusters and their strong heterogeneity couldn't lead to a global conclusion on the effect of the intervention.

Key words : Stepped wedge, Cluster Randomised Trials, Random effects models, Nonparametric methods, Within period analysis.

REMERCIEMENTS

En premier lieu, je souhaite remercier Laura Richert pour son attitude très positive concernant ma reconversion. Tu as instauré un climat favorable vis-à-vis de ce projet et a toujours été très disponible. De tout cela, je t'en remercie grandement.

Je souhaite également remercier tous les membres de l'équipe biostatistique de l'USMR pour leurs conseils et le temps consacré à l'étude OREANE. Je remercie en particulier Marion Kret, ma maître de stage, pour son encadrement dévoué. Je la remercie également de m'avoir appris à me débarrasser du pronom « on », si indésirable à la lecture. Je remercie également la méthodologiste de l'UMES Florence Saillour-Glenisson pour son implication tout au long de ce stage, ainsi que pour ses conseils. Merci à toutes les deux de m'avoir épaulé. J'ai également une pensée pour Julien Asselineau qui m'a conseillé de faire le DU « Méthodes statistiques de régression en épidémiologie » qui m'a permis de postuler à ce Master 2.

Je tiens également à remercier Monsieur Pierre Joly et Monsieur Robin Genuer, pour m'avoir reçu en entretien et conseillé avant le Master 2.

Enfin, je remercie infiniment ma compagne pour son soutien infaillible. Tu as été la source d'inspiration nécessaire à ma reconversion. Tu m'as apporté beaucoup de rigueur. Sans toi, je n'aurai eu ni la force ni le courage de me lancer.

TABLE DES MATIERES

1 INTRODUCTION 4

1.1 AVANT PROPOS 4
1.2 STRUCTURE D'ACCUEIL 4
1.2.1 Présentation générale de l'USMR 4
1.2.2 Présentation de l'équipe de l'USMR 5
1.2.3 Circuit d'une étude au sein de l'USMR 6
1.2.4 Rôles du biostatisticien au sein de l'USMR 8
1.3 LE SCHEMA STEPPED WEDGE : GENERALITES 8
1.3.1 Généralités méthodologiques 8
1.3.2 Autres types de design cross forward 10
1.3.3 Intérêts et limites du stepped wedge 10
1.4 PROBLEMATIQUE DE MON STAGE 12

2 OBJECTIFS DU STAGE 13

3 ENJEUX STATISTIQUES LIES AUX STEPPED WEDGE 14

3.1 LES DIFFERENTS NIVEAUX DE CORRELATION 14
3.1.1 Le cas transversal 14
3.1.2 Le cas de la cohorte 14
3.2 TAILLE D'ETUDE ET PUISSANCE 14
3.3 EFFET FRACTIONNE DE L'INTERVENTION 15
3.4 MODELISATION STATISTIQUE DES ESSAIS EN STEPPED WEDGE 15
3.4.1 Généralités 15
3.4.2 Les méthodes les plus retrouvées pour les stepped wedge 16
3.4.3 Les modèles statistiques : application au schéma transversal 16
3.4.3.1 Le modèle linéaire à effets aléatoires 17
3.4.3.2 Les extensions du modèle de base 17
3.4.3.3 Le modèle pour une variable réponse binaire 18
3.4.3.4 Les Equations d'Estimation Généralisées (GEE) 19
3.4.4 Les modèles statistiques : application au schéma cohorte 19
3.4.4.1 Le modèle linéaire à effets aléatoires 19
3.4.4.2 Le modèle pour une variable réponse binaire 20
3.5 LES APPROCHES NON PARAMETRIQUES 20
3.5.1 L'analyse intra-période (Thompson et al., 2018) 20
3.5.2 Les nouvelles méthodes non paramétriques 22

4 L'ETUDE OREANE 23

4.1 PRESENTATION DE L'ETUDE 23
4.1.1 Objectif principal 23

4.1.2 Les erreurs médicamenteuses ... *23*
4.1.3 Définition de l'intervention ... *24*
4.1.4 Schéma d'étude ... *24*
4.1.5 Particularités de l'étude OREANE ... *25*
4.2 MODELISATION STATISTIQUE DE L'ETUDE OREANE ... 26
4.2.1 Les variables du modèle sans covariable ... *26*
4.2.2 Modélisation sans covariable de l'étude OREANE ... *27*
4.2.3 Modélisation avec covariables de l'étude OREANE ... *28*
4.3 RESULTATS L'ETUDE OREANE ... 28
4.3.1 Descriptif des données ... *28*
4.3.1.1 Descriptif des établissements ... 28
4.3.1.2 Descriptif des actes médicaux réalisés par les IDE ... 29
4.3.1.3 Descriptif de la variable réponse ... 30
4.3.1.4 Descriptif de la variable réponse par centre ... 31
4.3.1.5 Les données manquantes de la variable réponse ... 33
4.3.2 Résultats de la modélisation ... *33*
4.3.2.1 Résultats du modèle sans covariable ... 33
4.3.2.2 Résultats du modèle avec covariables ... 34
4.3.3 Adéquation et discrimination ... *37*
4.3.3.1 Adéquation des modèles ... 37
4.3.3.2 Pouvoir discriminant des modèles ... 37
4.3.4 Application de l'analyse non paramétrique intra-période à l'étude OREANE ... *38*
5 DISCUSSION ET CONCLUSION ... 42
5.1 DISCUSSION ... 42
5.1.1 Le schéma stepped wedge ... *42*
5.1.2 Analyse statistique des stepped wedge ... *42*
5.1.3 Les taux d'erreur médicamenteuse ... *43*
5.1.4 Le faible nombre de clusters dans l'étude OREANE ... *44*
5.1.5 Effet temps aléatoire et/ou effet traitement aléatoire ... *44*
5.1.6 Discussion sur les résultats ... *45*
5.1.7 Les approches non paramétriques ... *46*
5.1.8 Les données manquantes de la variables réponse ... *46*
5.2 CONCLUSION ... 47
5.2.1 Conclusion du stage et perspectives ... *47*
5.2.2 Conclusion personnelle du stage ... *48*
6 REFERENCES BIBLIOGRAPHIQUES ... 49

LISTE DES FIGURES

Figure 1 : Organigramme de l'environnement dynamique de l'USMR du CHU de Bordeaux.5

Figure 2 : Organigramme hiérarchique et fonctionnel de l'USMR du CHU de Bordeaux en 2021. 6

Figure 3 : Intervention de l'USMR dans chaque étape du circuit d'une étude. 7

Figure 4 : Représentation du design stepped wedge. 9

Figure 5 : Représentation de plusieurs autres types de design. 10

Figure 6 : Représentation de la méthode non-paramétrique « analyse intra-période ». 21

Figure 7 : Représentation de la méthode non-paramétrique « méthode crossover ». 22

Figure 8 : Schéma de l'étude OREANE. 25

Figure 9 : Descriptif du taux d'erreur médicamenteuse des IDE par période d'observation.. 31

Figure 10 : Descriptif du taux d'erreur médicamenteuse des IDE par période d'observation par centre. 32

Figure 11 : Représentation des pouvoirs discriminants des modèles sans et avec covariables de l'étude OREANE. 38

Figure 12 : Présentation des taux d'erreur médicamenteuse avant et après l'intervention pour les données regroupées de la période deux et trois. 39

LISTE DES TABLEAUX

Tableau 1 : Participation des IDE aux différentes formes de l'intervention selon le centre.... 28

Tableau 2 : Descriptifs des covariables d'ajustement dans le cadre de l'analyse de l'étude OREANE..... 30

Tableau 3 : Représentation des données manquantes pour la variable « erreur médicamenteuse » par centre et par période d'observation. 33

Tableau 4 : Présentation du modèle initial d'OREANE : étude de l'effet de l'intervention globale sur le taux d'erreur médicamenteuse des IDE dans les unités de réanimation néonatale..... 34

Tableau 5 : Présentation du modèle final d'OREANE : étude de l'effet de l'intervention globale sur le taux d'erreur médicamenteuse des IDE dans les unités de réanimation néonatale avec covariables d'ajustement. 34

Tableau 6 : Résumé de la variable réponse de l'étude OREANE en fonction de la période et du statut vis-à-vis de l'intervention..... 39

Tableau 7 : Application de l'analyse intra-période aux données de l'étude OREANE. 40

Tableau 8 : Possibilités de permutations dans l'étude OREANE avec trois clusters (3! = 6). . 41

ABREVIATIONS

AIC : Akaike information criterion
AMM : Autorisation de Mise sur le Marché
AOI : Appel d'Offre Interne
AUC : Area Under Curve
BLUP : Best Linear Unbiaised Prediction
CAC : Cluster Auto Correlation
CHU : Centre Hospitalier Universitaire
CIC-EC : Centre d'Investigation Clinique de Bordeaux - Epidémiologie Clinique
CRT : Cluster Randomised Trial
DRCI : Direction de la Recherche Clinique et de l'Innovation
eCRF : electronic Case Report Form
GEE : Generalized Estimating Equation
GLMM : Generalized Linear Mixed Model
H2020 : Horizon 2020
IAC : Individual Auto Correlation
ICC : Intra Class Correlation
IDE : Infirmiers Diplômés d'Etat
IDERC : Infirmiers Diplômés d'Etat de Recherche Clinique
LMM : Linear Mixed Model
MAR : Missing At Random
OR : Odds Ratio
PAS : Plan d'Analyse Statistique
PHRCI : Programme Hospitalier de Recherche Clinique Interrégionaux
PHRC-K : Programme Hospitalier de Recherche Clinique en cancérologie
PHRC-N : Programme Hospitalier de Recherche Clinique Nationaux
PREPS : Programme de recherche sur la performance du système des soins
PRME : Programme de Recherche Médico-Economique
RASta : Rapport d'Analyse Statistique
ROC : Receiver Operating Characteristic
UMES : Unité de Méthodes d'Evaluation en Santé
USIN : Unité de Soins Intensifs Néonatale
USMR : Unité de Soutien Méthodologique à la Recherche clinique et épidémiologique

1 INTRODUCTION

1.1 Avant propos

La pandémie actuelle de Covid-19 a érigé les domaines de l'épidémiologie et des biostatistiques au premier plan de l'actualité. Au sein même de mon entourage, du plus proche au plus éloigné, tous plus ou moins se sont laissé tenter en faisant part aux autres de leurs commentaires ou de leur analyse de la situation parfois très tranchée. Force est de constater que le domaine des statistiques ne fait pas du tout l'unanimité dans la population générale (même si cela reste à prouver par une étude randomisée). Cette citation de Mark Twain en est une bonne illustration à mon sens : "Il y a trois sortes de mensonges : les mensonges, les sacrés mensonges et les statistiques". Faire un Master 2 de Biostatistique dans ce contexte pourrait alors créer un sentiment d'isolement plus important que le confinement lui-même. Cependant, les critiques à l'encontre de ce domaine ne sont parfois pas dénuées de bon sens, tant la statistique peut parfois amener à de fausses conclusions lorsqu'elle n'est pas utilisée à bon escient. J'y trouve alors une motivation constante pour progresser et aller plus loin.

Au cours de mon stage de Master 2 de Biostatistique réalisé à l'USMR, je me suis intéressé à la modélisation statistique des études en stepped wedge.

Ayant exercé le métier de data manager au sein de la recherche clinique pendant plus de trois ans à l'USMR, j'ai souhaité entamer une reconversion professionnelle pour devenir biostatisticien. C'est donc tout naturellement que je réalise mon stage dans cette unité. Ce sujet de stage m'a particulièrement intéressé car il m'a apporté une expertise dans l'analyse statistique d'un projet de recherche ainsi que des connaissances dans un schéma d'étude innovant.

1.2 Structure d'accueil

1.2.1 Présentation générale de l'USMR

Créée le 1er mai 1999, l'Unité de Soutien Méthodologique à la Recherche clinique et épidémiologique ou USMR, est une plate-forme épidémiologique, biostatistique et informatique impliquée dans les études de recherche clinique initiées majoritairement par le Centre Hospitalier Universitaire (CHU) de Bordeaux *(Figure 1).*

Les missions principales de l'USMR sont : la conception méthodologique, le recueil et le traitement de données, l'analyse des données et l'interprétation des résultats de projets de

recherches internes aux CHU de Bordeaux (AOI), interrégionaux (PHRCI), nationaux (PHRC-N, PHRC-K, PRME, PREPS) et européens (H2020). L'USMR collabore avec l'Unité des Méthodes d'Evaluation en Santé (UMES) sous la direction du Professeur Louis-Rachid Salmi dans le cadre de certains projets. La mission principale de l'UMES est d'apporter une contribution à l'appui méthodologique aux projets de recherche évaluative. Dans ces projets en collaboration, l'UMES est en charge des aspects méthodologiques, et l'USMR des aspects statistiques et de gestion des données.

L'USMR prend en charge plus de 160 projets provenant de différents pôles du CHU de Bordeaux et réalise une centaine de consultations méthodologiques chaque année. Depuis sa création, l'unité a contribué à la publication de nombreux articles scientifiques mais aussi à l'établissement et au fonctionnement de structures de recherche clinique de niveau international, impliquant le CHU et ses différents partenaires : Centre d'Investigation Clinique de Bordeaux dont le module Epidémiologie Clinique (CIC-EC 14-01), Centre de Traitement des Données cancer, ou encore la plateforme académique EUCLID/FCRIN pour des essais cliniques internationaux *(Figure 1)*.

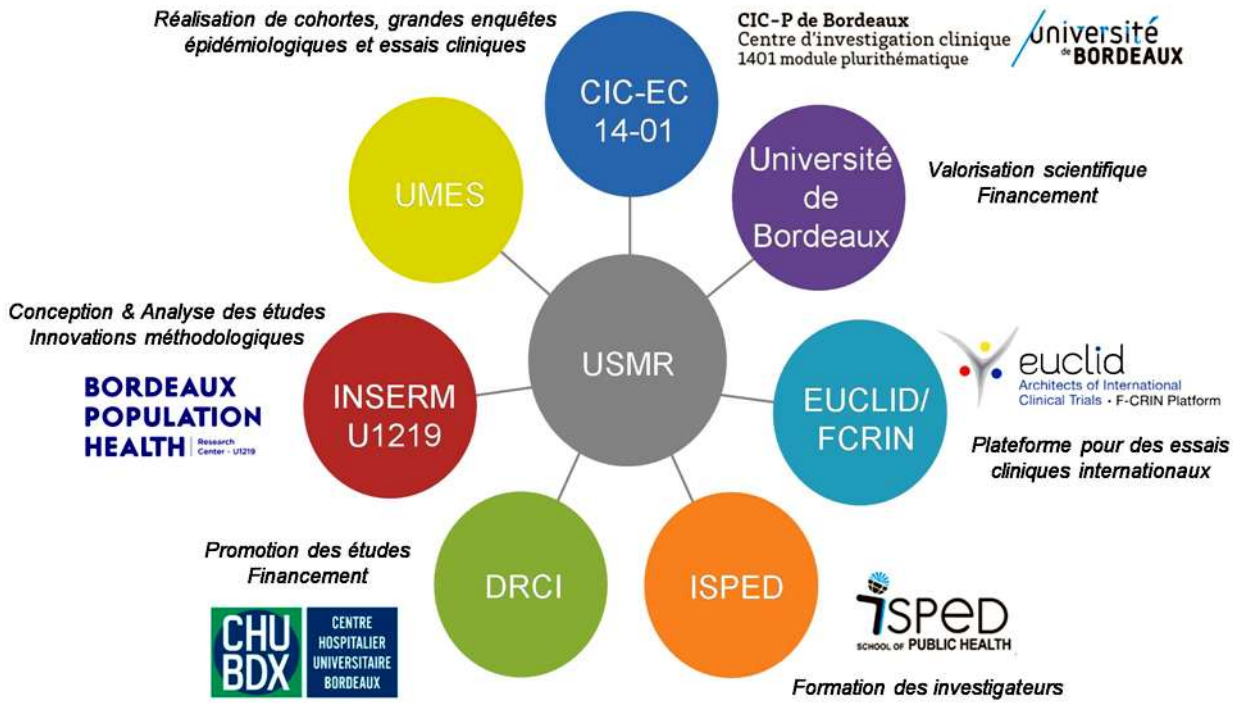


Figure 1 : Organigramme de l'environnement dynamique de l'USMR du CHU de Bordeaux.

1.2.2 Présentation de l'équipe de l'USMR

En 2021, l'équipe de l'USMR regroupe plus d'une vingtaine de membres incluant une responsable de l'unité, un responsable adjoint, une secrétaire, une responsable management

de la qualité et des risques, sept méthodologistes et plusieurs internes en médecine, ainsi que huit biostatisticiens, huit gestionnaires de données ou data managers et un informaticien *(Figure 2).*

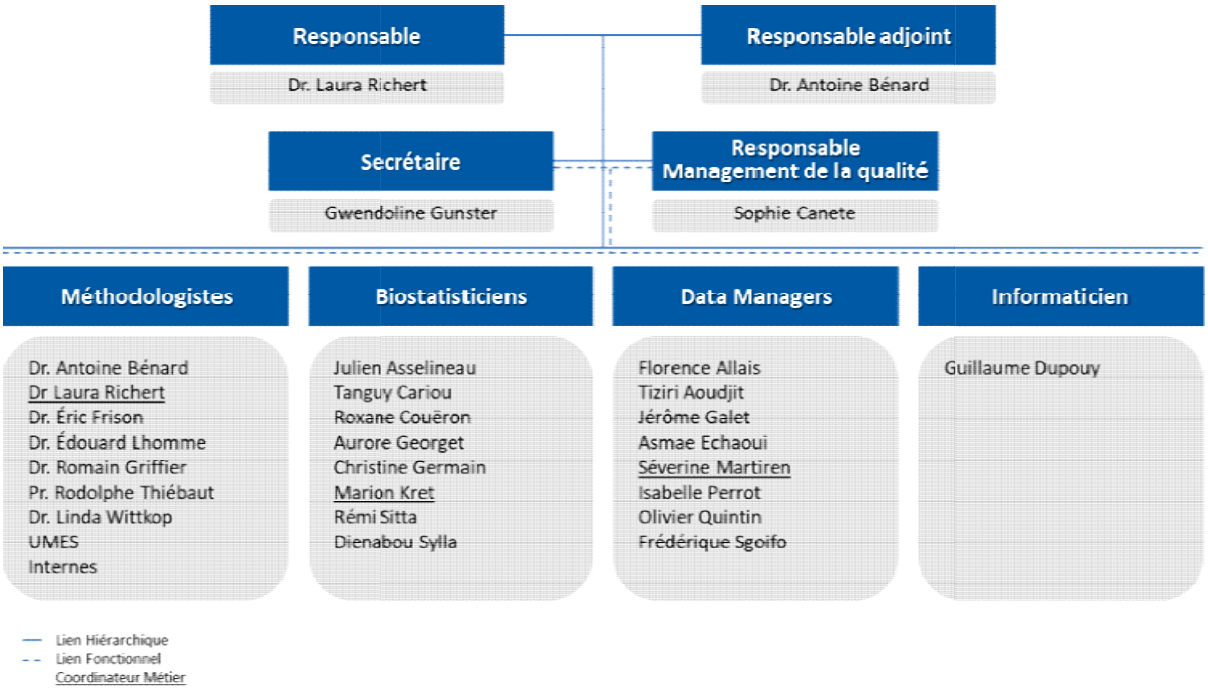


Figure 2 : Organigramme hiérarchique et fonctionnel de l'USMR du CHU de Bordeaux en 2021.

A titre personnel, je fais partie de cette unité depuis Avril 2017 en tant que data manager, et j'effectue le Master 2 de Biostatistique dans le cadre d'une reconversion professionnelle.

1.2.3 Circuit d'une étude au sein de l'USMR

Le circuit d'une étude, menée par une équipe investigatrice, comprend plusieurs étapes clés au cours desquelles l'aide méthodologique apportée par l'USMR s'étend de la soumission et la mise en place du projet jusqu'à la valorisation scientifique des résultats. Ces étapes sont regroupées au sein de quatre processus *(Figure 3) :*

- Le processus de conception : La consultation méthodologique représente la première étape du circuit d'une étude regroupant le traitement des aspects principaux de la recherche nécessaire à la réalisation d'un projet. L'accent est mis sur la conception méthodologique et le traitement statistique en vue de la rédaction du protocole de recherche : aide à la clarification de la question de l'étude, détermination des composants du schéma de l'étude et des méthodes statistiques, détermination de la taille de l'étude et aide à la rédaction des documents.
- Le processus de mise en place : Après obtention du financement, une réunion de lancement est programmée et assure le début de la mise en place du projet. La mise

en place d'une étude regroupe la finalisation de la conception du protocole, la vérification de sa faisabilité auprès des sites cliniques, la préparation de la liste des correspondants, la rédaction des documents nécessaires à l'étude (cahier d'observations, cahier des charges informatique…), la préparation de la base de données informatisée pour la saisie et le traitement des données recueillies pour l'étude et la préparation de la liste de randomisation dans le cas des études randomisées.

- Le processus de suivi : Le suivi d'une étude consiste en premier lieu à vérifier que cette dernière se déroule conformément au protocole et aux référentiels en collaboration avec le coordinateur. C'est lors de cette étape que le contrôle de la base de données est réalisé sous la responsabilité du data-manager. Ce processus se termine par la rédaction d'un plan d'analyse statistique et par la réalisation d'une revue des données sous la responsabilité du biostatisticien.
- Le processus d'analyse et de valorisation : La dernière étape du circuit d'une étude comprend l'analyse et la valorisation des résultats via la rédaction du rapport d'analyse statistique et la publication scientifique.

Le rôle clé joué par l'USMR dans l'ensemble de ces étapes du circuit d'une étude est effectué en collaboration constante avec la DRCI (Direction de la Recherche Clinique et de l'Innovation) du CHU de Bordeaux.

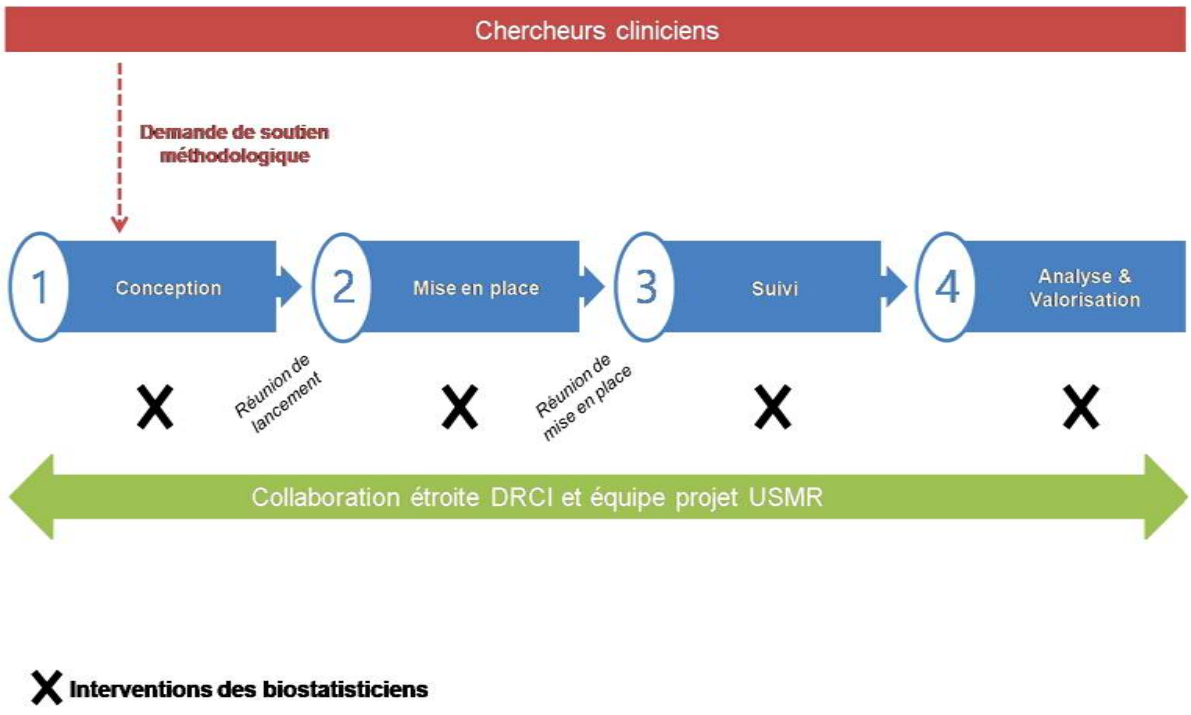


Figure 3 : Intervention de l'USMR dans chaque étape du circuit d'une étude.

1.2.4 Rôles du biostatisticien au sein de l'USMR

Les biostatisticiens interviennent dans l'ensemble des phases du circuit d'une étude, et notamment dans la phase de conception en apportant une réflexion poussée sur les choix méthodologiques alimentée par des recherches bibliographiques, l'estimation de la taille de l'étude, ou encore en participant à la rédaction des aspects statistiques du protocole *(Figure 3)*. En outre, ils participent à la phase de mise en place de l'étude et de planification afin de définir et d'implémenter la liste de randomisation dans le cas d'essais randomisés, mais aussi assurent la relecture du cahier d'observation, dédié au recueil de l'ensemble des données requises par le protocole de l'étude. Par ailleurs, les biostatisticiens sont en charge du bon déroulement de la recherche sur le plan méthodologique avec le suivi et la gestion des non respects du protocole (défaut d'inclusion, inclusions à tort, déviations au protocole,...), la rédaction du plan d'analyse statistique (PAS), la revue des données, et la réalisation d'analyses statistiques nécessaires pour les réunions du conseil scientifique et du comité indépendant de surveillance. Enfin, lors de la phase finale du circuit d'une étude, ils effectuent les analyses statistiques des données, l'interprétation statistique des résultats et la rédaction du rapport d'analyse statistique (RASta). En collaboration avec les méthodologistes et l'équipe investigatrice de l'étude, les résultats sont discutés, rédigés dans un rapport de fin d'étude et valorisés sous forme de communications orales et/ou articles pour l'ensemble de la communauté scientifique.

Des activités transversales sont également réalisées par les biostatisticiens au sein de l'USMR comme la rédaction et la révision des documents de qualité, l'enseignement de la méthodologie de la recherche clinique, le développement d'innovations méthodologiques, une veille bibliographique des méthodes statistiques ou encore l'encadrement statistique des internes de santé publique et des stagiaires.

1.3 Le schéma stepped wedge : généralités

1.3.1 Généralités méthodologiques

Le schéma d'étude en stepped wedge fait partie des « Clusters Randomised Trials » ou CRT. L'unité d'intérêt pour les CRT est le « cluster », soit un groupe identifiable et préexistant tel qu'un centre hospitalier, une école ou encore une entreprise. C'est donc au niveau de ce groupe que l'on souhaite apporter des conclusions sur l'effet d'une intervention ou bien la mise en place d'un dispositif par exemple. D'autre part, les CRT se caractérisent par le fait que la randomisation s'effectue au niveau du cluster et non au niveau individuel. Le schéma d'étude en stepped wedge présente la particularité d'introduire progressivement une intervention au sein des clusters. Il s'agit d'un type de design « crossover » au cours duquel les différents clusters changent d'état à des temps différents dans un sens unique (cross forward en anglais), habituellement du contrôle vers l'intervention, ce qui confère à ce

schéma son aspect en marche d'escalier ou « step » en anglais. Au début de l'essai, aucun cluster ne reçoit l'intervention et ce durant une première période de temps définie appelée « période d'observation (ou de mesure) ». Par la suite, l'intervention est implémentée de manière séquentielle dans chacun des clusters selon un ordre déterminé par la randomisation, jusqu'à ce que tous les clusters soient exposés à la dernière période d'observation. Une période d'observation contenant un ou plusieurs clusters en condition d'intervention est également appelée « séquence ». Le nombre de périodes d'observation est donc égal au nombre de séquences plus le nombre de périodes en condition de contrôle (habituellement une seule période dans le schéma stepped wedge classique). Dans certains cas, une période de transition peut être nécessaire à l'implémentation de l'intervention, celle-ci est généralement de la même durée qu'une période d'observation.

Ce schéma d'étude est adapté dans l'évaluation de l'effet d'une intervention au niveau d'un système de soins, en particulier lorsque l'introduction de l'intervention implique un effet irréversible. Dans ce cas, un retour du cluster exposé à la condition de contrôle n'est pas envisageable car l'effet de l'intervention ne peut être annihilé (habituellement appelé « carry over effect » en anglais). Dans le cas contraire où un retour en arrière est possible, le schéma de type crossover classique est plus adapté.

La figure 4, ci-dessous, est une illustration de ce schéma d'étude (Hemming et al., 2015).



Figure 4 : Représentation du design stepped wedge. A gauche : le schéma d'étude classique, à droite : le schéma incluant une période de transition.

A ce jour, il existe une grande diversité de types de schéma stepped wedge regroupés en deux principaux : le type transversal et le type cohorte.

Dans le cas du schéma transversal, l'échantillon de sujets observés dans un cluster donné est renouvelé à chaque période d'observation.

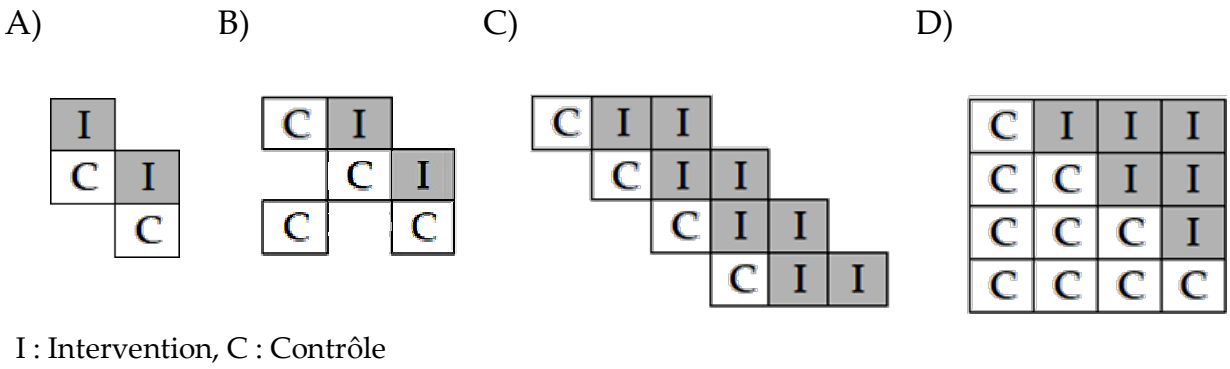
Dans le cas du schéma en cohorte fermée, l'échantillon de sujets observés dans un cluster donné est identique à chaque période d'observation. L'aspect longitudinal de la cohorte dû

aux données répétées des sujets implique un niveau de corrélation supplémentaire à prendre en considération par rapport au cas transversal. En effet, un sujet donné peut être observé autant de fois qu'il y a de périodes prévues dans l'étude. Dans le cas particulier de la cohorte ouverte, certains sujets peuvent arriver en cours d'étude et d'autres peuvent sortir.

1.3.2 Autres types de design cross forward

Il existe d'autres types de design cross forward. Le « dog leg », par exemple, est un schéma incomplet qui nécessite moins de participant que le stepped wedge car tous les clusters ne sont pas observés à toutes les périodes, ce qui confère à ce dernier son aspect en patte de chien (Hooper and Bourke, 2015). Les schémas incomplets sont parfois utilisés pour des raisons de faisabilités. Cependant, ils sont moins puissants que le schéma stepped wedge car ils comprennent moins de périodes de mesures et ne permettent pas d'ajuster correctement l'effet de l'intervention sur une tendance temporelle (Hemming et al., 2015b).

Par ailleurs, il existe également des variantes du stepped wedge classique, comme le schéma stepped wedge hybride qui est un mélange entre le schéma parallèle et le schéma stepped wedge (Thompson et al., 2017), qui comporte un cluster contrôle qui ne reçoit pas l'intervention durant toute la période de recueil. La figure 5, ci-dessous, est une illustration de ces quelques variantes.



I : Intervention, C : Contrôle

Figure 5 : Représentation de plusieurs autres types de design : A) le schéma dog leg classique, B) le schéma dog leg incluant une période « baseline », C) le schéma incomplet « un avant et deux après » et D) le schéma hybride.

1.3.3 Intérêts et limites du stepped wedge

Les essais contrôlés randomisés utilisant un schéma en stepped wedge sont récents. La première étude suivant ce type de schéma a en effet été réalisée en Gambie en 1986 et portait sur la vaccination contre l'hépatite B. Depuis, l'utilisation des essais en stepped wedge s'est rapidement répandue, tout comme les articles concernant la méthodologie et l'analyse

statistique dans ce domaine (Barker et al., 2016) car ils proposent un schéma en cluster adapté à des problématiques de recherche nouvelles (recherche sur les services de santé), et ce pour plusieurs raisons (Mdege et al., 2011) :

- Ils constituent un avantage dans le cas où tous les centres ne peuvent pas, pour des contraintes logistiques, bénéficier de l'intervention de manière simultanée.
- Tous les clusters participants à l'étude bénéficient de l'intervention supposée favorable ; ce qui a pour conséquence de faciliter leurs recrutements. Cela permet par ailleurs de s'affranchir d'un éventuel non respect de la clause d'ambivalence.
- Dans certains cas, le tirage au sort dans le groupe de traitement ou le groupe contrôle peut être considéré éthiquement non acceptable. Le schéma stepped wedge constitue alors une alternative favorable car tous les clusters y reçoivent l'intervention.
- Ce schéma nécessite moins de clusters que le schéma parallèle car en effet, chacun des clusters est à la fois non exposé puis exposé selon le temps considéré.
- Enfin, sa réalisation ne nécessite pas l'obtention du consentement éclairé des participants.

Les essais en stepped wedge doivent toutefois être pensés avec précaution car ils présentent des difficultés par ailleurs :

- Le respect du calendrier du schéma peut s'avérer difficile.
- Le temps nécessaire pour la conduite d'un stepped wedge est plus important que pour le schéma parallèle.
- Les clusters en condition de contrôle sont plus nombreux au début de l'étude et les clusters exposés sont plus nombreux en fin d'étude. Par conséquent, le temps calendaire peut agir comme facteur de confusion de l'évènement, et un ajustement sur le temps ou la période est donc nécessaire (Hemming et al., 2017).
- L'effet complet de l'intervention doit, dans l'idéal, être atteint durant la période à laquelle l'intervention a été introduite. Plus le retard de l'effet de l'intervention est grand, plus important est le manque de puissance. Le schéma n'est donc pas adapté lorsqu'il existe une interaction sur le long terme entre le temps et l'intervention.
- Ce schéma peut être vécu comme lourd et invasif dans le cas du recueil de données répétées au niveau individuel (spécifique au cas cohorte).

De manière plus générale et non spécifique aux schémas en stepped wedge, il peut exister dans les essais en cluster un risque de contamination de l'effet de l'intervention aux clusters non exposés.

1.4 Problématique de mon stage

La littérature sur les stepped wedge est récente et les connaissances sur ce sujet nouvelles. Le premier enjeu de mon stage était de faire la synthèse des différentes méthodes d'analyse statistique des schémas en stepped wedge et d'identifier les pistes de recherche et d'explorations nouvelles sur le sujet. Le second enjeu était d'explorer les différentes pistes d'analyse statistique d'une étude utilisant un schéma en stepped wedge de type cohorte : l'étude OREANE.

L'étude OREANE (financée par l'appel d'offre du PREPS) s'intéresse à l'effet d'une intervention complexe sur le taux d'erreur médicamenteuse des Infirmiers Diplômés d'Etat (IDE) en unités de réanimation néonatale, connues comme présentant des forts taux d'erreur (principalement erreurs d'administration). Etant donné la nature de l'intervention qui s'implémente de manière décalée au niveau du centre hospitalier, le schéma de l'étude OREANE est bien un stepped wedge. Toutefois, la méconnaissance de ce type de schéma à la construction du projet a impliqué que plusieurs éléments méthodologiques n'ont pas été prévus selon les recommandations. Tout l'enjeu de mon stage a donc été de bien comprendre les méthodes d'analyse des stepped wedge afin de pouvoir les appliquer à l'étude OREANE, tout en prenant en compte ces contraintes méthodologiques.

La première étape dans la réalisation de mon stage a été de faire une revue bibliographique concernant ce schéma d'étude et plus spécifiquement sur les méthodes d'analyse statistique qui y sont associées. Par la suite, j'ai pu découvrir la base de données de l'étude OREANE et commencer les analyses sur les logiciels SAS et R. Durant tout le stage, j'ai également pu participer aux différentes discussions avec l'équipe investigatrice et l'équipe projet ainsi qu'aux différentes réunions consacrées à cette étude avec les autres biostatisticiens de l'USMR.

2 OBJECTIFS DU STAGE

Les objectifs proposés au cours de ce stage sont :

- Décrire et définir les indications des différents modèles statistiques les plus retrouvés dans la littérature à ce jour en fonction des différents types de stepped wedge (transversal et cohorte). Ce travail se base sur une revue approfondie de la littérature concernant l'analyse des stepped wedge et des études randomisées en cluster de manière plus générale.
- Identifier le modèle linéaire mixte généralisé (GLMM) le plus adapté aux données de l'étude OREANE. Le modèle final sera retenu selon des critères de convergences et d'adéquations.
- Explorer de nouvelles méthodes non paramétriques ayant vu le jour récemment pour l'analyse des stepped wedge (Kennedy-Shaffer et al., 2020). La première étape sera une description de ces méthodes pour savoir si l'une d'entre elles pourrait s'appliquer à l'étude OREANE et le cas échéant, une mise en parallèle des résultats pourra être effectuée avec ceux obtenus en utilisant l'approche GLMM.

3 ENJEUX STATISTIQUES LIES AUX STEPPED WEDGE

3.1 Les différents niveaux de corrélation

Les essais en stepped wedge comportent des données répétées au niveau du cluster et/ou au niveau individuel. Ceci implique de prendre en considération ces différents niveaux de corrélation. Considérons : Y_{ijk}, le niveau de réponse individuel k (k=1,…,N) du cluster i (i=1,…,I) à la période j (j=1,…,T).

3.1.1 Le cas transversal

Dans le cas transversal, les données sont répétées au niveau du cluster ce qui implique la nécessité de prendre en compte cette corrélation intra-groupe. D'une part, l' « Intra Class Correlation » (ICC) représente la corrélation entre les mesures de deux individus différents au sein d'un même cluster, à la même période d'observation ($Corr(Y_{ijk},Y_{ijk'})$). On peut également la retrouver sous l'appellation de corrélation intra-période. D'autre part, la « Cluster Auto Correlation » (CAC) ou corrélation inter-période caractérise la corrélation entre les mesures de deux individus différents au sein d'un même cluster, mais à des périodes d'observations différentes ($Corr(Y_{ijk},Y_{ij'k'})$).

3.1.2 Le cas de la cohorte

Dans le cas de la cohorte, on retrouve l'ICC et le CAC car les mesures du cluster sont également répétées dans le temps. Cependant, il est nécessaire de tenir compte en plus du niveau de corrélation intra-sujet (ou IAC pour « Individual Auto Correlation ») qui est défini comme la corrélation entre deux mesures d'un même individu à des temps différents ($Corr(Y_{ijk},Y_{ij'k})$).

3.2 Taille d'étude et puissance

Toutes les études en cluster doivent prendre en compte la corrélation entre les individus d'un même cluster. C'est pourquoi une étude en cluster parallèle nécessite plus de sujets qu'une étude randomisée au niveau individuel. Le calcul de la taille d'étude dépend d'une part de la nature du stepped wedge (transversal ou de type cohorte), mais également des hypothèses sur la structure de corrélation. L'approche standard suppose que deux observations d'un même cluster ont une corrélation constante : la corrélation intra-cluster ICC (mais ce n'est pas toujours le cas). L'inflation de la taille d'étude dépend donc d'un

facteur multiplicatif (couramment appelé « design effect » en anglais) qui dépend de cette corrélation.

Cependant, dans les études stepped wedge avec de multiples périodes de temps, une structure de corrélation de type échangeable, qui suppose que la corrélation entre deux observations du même cluster est identique quel que soit le temps qui sépare les observations, n'est pas la plus adaptée. En effet, il est souvent nécessaire de prendre en compte une corrélation inter-périodes au niveau du cluster (« Cluster Auto Correlation » ou CAC). On peut également supposer que la CAC soit décroissante au cours du temps, de façon exponentielle par exemple, en introduisant une composante autorégressive dans la modélisation (Hemming et al., 2020).

D'autres paramètres peuvent également diminuer la puissance et, par conséquent, augmenter le nombre de sujets nécessaires comme des tailles de clusters différentes, un effet fractionné de l'intervention, (*voir paragraphe ci-dessous*) ou encore la présence de données corrélées au niveau individuel : IAC.

3.3 Effet fractionné de l'intervention

Dans le schéma stepped wedge classique, on suppose que l'effet complet de l'intervention se produit dès que le cluster passe de la condition de contrôle à intervention. En pratique, il arrive que l'effet de l'intervention ne soit pas immédiat et dépende du temps. Par exemple, l'effet pourrait être de 50% à la première période d'observation où le cluster reçoit l'intervention, de 75% au second et de 100% au troisième. Ce phénomène se nomme « effet fractionné de l'intervention » et il en résulte un manque de puissance si les clusters ne sont pas observés suffisamment longtemps pour que l'effet de l'intervention soit pleinement atteint. Cette perte de puissance peut être partiellement récupérée grâce à l'ajout de périodes d'observations à la fin de l'étude. D'autre part, les conséquences indésirables occasionnées par cet effet fractionné peuvent être anticipées en le prenant en considération dans le calcul de la taille d'étude (Voldal et al., 2020).

3.4 Modélisation statistique des essais en stepped wedge

3.4.1 Généralités

Une des premières méthodes d'analyse retrouvée dans la littérature est l'analyse intra-cluster pour estimer l'effet d'une intervention. Cependant, cette analyse est biaisée s'il existe un effet lié au temps, car ce dernier n'est pas pris en considération et agit par conséquent comme un facteur de confusion (Hussey and Hughes, 2007). Les modèles statistiques plus récents, comme les modèles à effets aléatoires par exemple, utilisent l'information intra et inter-cluster, ce qui permet d'éviter cet effet de confusion lié à des changements au cours du

temps. L'enjeu est d'utiliser des méthodes adaptées aux différents types de stepped wedge et suffisamment flexibles pour prendre en compte les spécificités de chaque projet.

3.4.2 Les méthodes les plus retrouvées pour les stepped wedge

Les méthodes utilisées dans l'analyse des stepped wedge dépendent de plusieurs facteurs comme la nature du schéma (transversal ou cohorte), la nature de la variable réponse recueillie ou encore le nombre de clusters présents dans l'étude. Une revue des méthodes statistiques utilisées a été faite (Barker et al., 2016). Celle-ci indique que la plupart des études en schéma stepped wedge utilise soit une approche mixte (LMM ou GLMM selon la nature de la variable réponse), soit l'approche des Equations d'Estimation Généralisées (GEE). La méthode utilisée pour les stepped wedge de type cohorte est presque exclusivement une approche mixte. En revanche, pour les stepped wedge transversaux, les deux approches sont retrouvées (GEE et mixte). L'utilisation de l'approche GEE dans le cas transversal peut se justifier par le fait que tous les clusters sont observés à tous les temps. Il n'y a donc pas de données manquantes comme il peut y en avoir dans l'approche longitudinale où les sujets sont observés à des temps différents et peuvent éventuellement quitter l'étude en cours. Dans le cas transversal, on ne peut donc pas privilégier l'approche mixte qui est plus robuste en cas de données manquantes.

Grâce à des simulations, une étude a comparé plusieurs méthodes pour les stepped wedge transversaux (LMM, GLMM et GEE) en faisant varier le nombre de clusters, le nombre de périodes d'observations et la corrélation intra-classe (ICC). Les comparaisons portaient sur le biais d'estimation, la puissance, l'erreur de type I ainsi que les problèmes de convergence. En conclusion, ils recommandent l'utilisation de l'approche GLMM lorsque la réponse est binaire et qu'il y a peu de clusters dans l'étude (peu étant défini comme trois). En outre, le nombre minimum de clusters conseillé pour un stepped wedge est de six pour trois séquences. En dehors de ce scénario, les méthodes GEE et GLMM sont comparables en termes de biais (Barker et al., 2017).

3.4.3 Les modèles statistiques : application au schéma transversal

Les modèles statistiques inhérents aux stepped wedge doivent permettre de répondre à deux problématiques importantes : la corrélation entre les individus d'un même cluster et le potentiel effet de confusion que peut avoir l'évolution naturelle des pratiques dans les centres, nécessitant un ajustement sur le temps.

3.4.3.1 Le modèle linéaire à effets aléatoires

Les modèles à effets aléatoires sont souvent utilisés pour prendre en compte la corrélation entre les individus du même cluster. Le modèle linéaire à effets mixtes de base proposé par Hussey et Hughes prend en considération I clusters, T temps et N individus par cluster par intervalle de temps (Hussey and Hughes, 2007). Soit Y_{ijk} la variable réponse correspondant à l'individu k au temps j du cluster i (i de 1, …, I ; j de 1, …, T ; k de 1, …, N) et Y_{ij} la moyenne du cluster i au temps j, ce qui définit :

$$\mu_{ij} = \mu + \alpha_{0i} + \beta_j + X_{ij}\theta$$

où α_{0i} est l'intercept aléatoire du cluster i tel que $\alpha_i \sim N(0, \sigma_0^2)$, β_j est l'effet fixe correspondant à l'intervalle de temps j, X_{ij} est une variable indicatrice de l'intervention du cluster i au temps j (1 pour intervention et 0 pour contrôle) et θ l'effet du traitement.

Le niveau de réponse individuel est modélisé tel que :

$$Y_{ijk} = \mu_{ij} + e_{ijk}$$

Avec $e_{ijk} \sim N(0, \sigma_e^2)$. On suppose l'indépendance entre e_{ijk} et α_i.

Dans ce modèle à intercept aléatoire, la corrélation intra-classe ou intra-cluster s'écrit : $ICC = \frac{{\sigma_0}^2}{{\sigma_0}^2 + {\sigma_e}^2}$ et cette quantité représente la corrélation existante entre deux individus d'un même cluster.

Ce modèle fait plusieurs hypothèses importantes. La corrélation intra-cluster est supposée constante dans le temps car il n'y a qu'un intercept aléatoire. Le temps dans ce modèle est dans les effets fixes, ce qui suppose une tendance temporelle commune à tous les clusters. De même, il comporte un terme unique pour l'effet du traitement, ce qui implique un effet commun de cette tendance à tous les temps de mesure (pas d'interaction avec le temps). En outre, le temps est catégoriel. Ce modèle implique également une structure de corrélation simple où la corrélation entre deux observations du même cluster est la même, indépendamment du traitement et du délai entre ces observations. Enfin, il suppose également que les données recueillies à des temps différents proviennent d'individus différents, ce qui est spécifique au cas transversal.

3.4.3.2 Les extensions du modèle de base

L'évolution de la connaissance des stepped wedge a permis d'introduire des extensions à ce modèle pour apporter plus de flexibilité, par exemple en ajoutant des effets aléatoires sur le traitement et/ou sur le temps comme illustré dans le modèle ci-dessous (Voldal et al., 2020).

$$\mu_{ij} = \mu + \alpha_{0i} + \beta_j + X_{ij}\theta + \alpha_{1ij} + \alpha_{2i} X_{ij}$$

où α_{1ij} est la pente aléatoire du cluster i tel que $\alpha_{1ij} \sim N(0, \sigma_1^2)$ et α_{2i} est l'effet du traitement aléatoire du cluster i tel que $\alpha_{2i} \sim N(0, \sigma_2^2)$. La corrélation entre l'intercept aléatoire α_i et l'effet du traitement aléatoire α_{2i} est σ_{20} et la corrélation entre d'une part α_{1ij} et α_{0i} et d'autre part α_{1ij} et α_{2i} sont toutes deux nulles. Enfin, la corrélation entre α_{1ij} et $\alpha_{1ij'}$ avec $j \neq j'$ est nulle.

Le niveau de réponse individuel est toujours :

$$Y_{ijk} = \mu_{ij} + e_{ijk}$$

avec $e_{ijk} \sim N(0, \sigma_e^2)$. On suppose l'indépendance mutuelle entre e_{ijk} et tous les effets aléatoires α_{0i}, α_{1ij} et α_{2i}.

Ce modèle est plus souple que le précédent, mais il peut arriver qu'il présente des problèmes de convergence car il contient plus d'effets aléatoires. Dans ce cas, un choix peut être nécessaire entre la pente aléatoire et l'effet du traitement aléatoire. Le choix dépend du contexte et des hypothèses spécifiques à chaque cas de figure. Il n'y a pas de recommandation pour aider dans ce choix sauf dans le cas d'un faible nombre de séquences, où il est conseillé d'introduire un effet temps aléatoire plutôt qu'un effet traitement aléatoire. Le choix des effets aléatoires est important car il peut changer les estimations des paramètres.

Si l'on considère qu'il n'y a pas d'effet du traitement aléatoire ($\sigma_2^2 = 0$), la quantité nommée ICC représente la corrélation intra-cluster au sein d'une même période. Cette nuance supplémentaire doit être ajoutée par rapport au modèle à intercept aléatoire en raison de la structure de covariance plus complète grâce à la présence d'un effet dépendant du temps spécifique au cluster. En effet, l'intercept aléatoire, ici, représente la variation du niveau de réponse initiale pour chaque cluster par rapport à μ. L'ICC est alors le ratio entre la variance inter-cluster et la variance totale : $ICC = \frac{{\sigma_0}^2 + {\sigma_1}^2}{{\sigma_0}^2 + {\sigma_1}^2 + {\sigma_e}^2}$.

L'auto-corrélation du cluster (CAC), est le ratio entre la corrélation interindividuelle au sein d'un même cluster à des temps différents et la corrélation interindividuelle au sein du cluster à la même période, d'où la relation : $CAC = \frac{{\sigma_0}^2}{{\sigma_0}^2 + {\sigma_1}^2}$.

3.4.3.3 Le modèle pour une variable réponse binaire

Dans le cas où la variable réponse Y_{ijk} est binaire suivant une loi binomiale [0 ;1], le modèle statistique peut être obtenu en utilisant les modèles linéaires généralisés à effets mixtes (approche GLMM). Par exemple, un modèle à intercept aléatoire s'écrit :

$$Logit[Pr(Y_{ijk}=1)] = \mu + \alpha_{0i} + \beta_j + X_{ij}\theta$$

où α_{0i} est l'intercept aléatoire du cluster i tel que $\alpha_i \sim N(0, \sigma_0^2)$. L'approche GLMM permet de traiter tous les modèles appartenant à la famille exponentielle.

3.4.3.4 Les Equations d'Estimation Généralisées (GEE)

L'approche GEE permet de déterminer une matrice de covariance de travail, c'est pourquoi elle est en pratique souvent utilisée dans les études stepped wedge transversales. Cependant, cette approche n'a pas été utilisée dans le cadre de ce stage qui s'intéresse principalement à l'analyse de l'étude OREANE, qui est un stepped wedge de type cohorte présentant des données manquantes. Cette méthode est définie comme suit (Li et al., 2018) :

$$g(\mu_{ijk}) = \beta_j + X_{ij}\theta$$

avec μ_{ijk} comme réponse marginale moyenne et θ, l'effet marginal de l'intervention dans l'échelle de la fonction de lien g().

La matrice de covariance de travail est habituellement notée V_i et peut être définie sous différentes formes (échangeable, à symétrie de composition ou autorégressive).

3.4.4 Les modèles statistiques : application au schéma cohorte

3.4.4.1 Le modèle linéaire à effets aléatoires

Le premier modèle linéaire à effets aléatoires s'appliquant aux stepped wedge de type cohorte a été décrit récemment (Li et al., 2018). Il s'agit d'un modèle multi-niveaux permettant de prendre en compte les différents niveaux de corrélations (au niveau du cluster et au niveau individuel). Le modèle proposé s'écrit :

$$Y_{ijk} = \mu + \beta_j + X_{ij}\theta + \alpha_{0i} + \alpha_{1ij} + \alpha_{2ik} + e_{ijk}$$

où α_{2ik} est l'intercept aléatoire du sujet k inclus dans le cluster i tel que $\alpha_{2ik} \sim N(0, \sigma_2^2)$. Les effets aléatoires et l'erreur individuelle (α_{0i}, α_{1ij}, α_{2ik} et e_{ijk}) sont supposés indépendants les uns des autres et représentent la variance totale du niveau de réponse individuel Y_{ijk} obtenue par la somme des composants $\sigma_0^2 + \sigma_1^2 + \sigma_2^2 + \sigma_e^2 = \sigma^2$. Grâce à ce modèle, on définit la corrélation intra-cluster telle que : $ICC = \frac{\sigma_0^2+\sigma_1^2}{\sigma^2}$,

l'auto-corrélation du cluster : $CAC = \frac{\sigma_0^2}{\sigma^2}$,

et enfin, l'auto-corrélation individuelle telle que : $IAC = \frac{\sigma_0^2+\sigma_2^2}{\sigma^2}$ qui correspond à la corrélation entre deux individus différents du même cluster à des temps différents.

Certaines extensions peuvent être apportées à ce modèle comme par exemple en ajoutant d'autres effets aléatoires au niveau du cluster ou bien au niveau individuel. Cependant, un modèle sur-paramétré peut présenter des problèmes de convergence, c'est pourquoi le choix des effets aléatoires doit être basé sur des hypothèses plausibles.

3.4.4.2 Le modèle pour une variable réponse binaire

Dans le cas où la variable réponse Y_{ijk} est binaire suivant une loi binomiale [0 ;1], le modèle statistique peut être obtenu comme précédemment en utilisant les modèles linéaires généralisés à effets mixtes (approche GLMM). Par exemple, un modèle à intercept aléatoire s'écrit :

$$\text{Logit}[\Pr(Y_{ijk}=1)] = \mu + \alpha_{0i} + \alpha_{1ik} + \beta_j + X_{ij}\theta$$

où α_{0i} est l'intercept aléatoire du cluster i tel que $\alpha_i \sim N(0, \sigma_0^2)$ et l'on retrouve l'intercept aléatoire du niveau individuel k tel que $\alpha_{1ik} \sim N(0, \sigma_2^2)$. Comme pour l'exemple précédent, ce modèle peut être enrichi par l'ajout d'effets aléatoires.

3.5 Les approches non paramétriques

Pour les essais en stepped wedge comportant un nombre relativement restreint de clusters, l'inférence asymptotique basée sur l'hypothèse d'effets aléatoires normalement distribués - comme cela est souvent fait pour l'analyse à l'aide de modèles à effets mixtes - peut entraîner une surestimation de l'erreur de type I ainsi qu'une mauvaise couverture de l'intervalle de confiance. Certaines méthodes non paramétriques pour l'analyse des stepped wedge ont été récemment proposées et constituent des alternatives robustes pour les essais avec peu de clusters.

3.5.1 L'analyse intra-période (Thompson et al., 2018)

Cette méthode s'intéresse à l'estimation de l'effet de l'intervention au sein de périodes d'observation et est également connue sous le nom de comparaison verticale. Pour chaque période où il y a des clusters dans le contrôle et dans l'intervention, une estimation de l'effet spécifique de la période est calculée en comparant le résultat moyen des clusters en condition d'intervention au résultat moyen des clusters en condition de contrôle.

La figure 6, ci-dessous, est une illustration de cette approche.

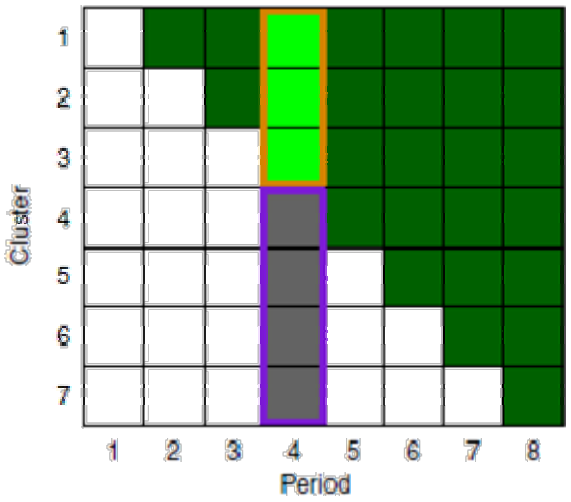


Figure 6 : Représentation de la méthode non-paramétrique « analyse intra-période » (Kennedy-Shaffer et al., 2020).

La méthode se décompose en trois étapes :

- La première est de faire un résumé (moyenne, écart-type et effectif) de la variable réponse Y_{ij} du cluster i à la période j dans les conditions de contrôle et d'exposition à l'intervention. Dans le cas où la variable réponse est binaire, le résumé doit se faire dans l'échelle du logit : $Y_{ij}^{*}=\log\left(\frac{p_{ij}}{1-p_{ij}}\right)$
- La seconde est d'estimer les effets de l'intervention $\widehat{\theta_j}$ comme étant la différence entre la moyenne des clusters dans l'intervention avec celle des clusters dans le contrôle au sein de la période j :

$\widehat{\theta_j}=\frac{1}{c_{1j}}*\sum_{i:X_{ij}=1} Y_{ij} - \frac{1}{c_{0j}}*\sum_{i:X_{ij}=0} Y_{ij}$

où c1j et c0j sont respectivement les clusters en condition d'intervention et de contrôle.

- La troisième étape consiste à combiner les estimations des effets propres à chaque période en utilisant une moyenne pondérée égale à l'inverse de la variance pour estimer l'effet global de l'intervention $\widehat{\theta}$:

$\widehat{\theta}=\sum_{j=1}^{T}\frac{w_j}{w}\widehat{\theta_j}$, avec $w=\sum_{j=1}^{T} w_j$ où w_j est l'inverse de la variance de la réponse au sein de la période j.

La p-value ainsi que l'intervalle de confiance peuvent être obtenus en utilisant un test de permutation, qui ne se base plus sur une distribution théorique mais empirique, car construite directement à partir des données. Cette approche ne nécessite aucune hypothèse sur la corrélation entre les effets de chaque période. Les clusters sont supposés indépendants les uns des autres au sein de chaque période. Egalement, l'échangeabilité des observations au sein de chaque cluster par période est une hypothèse nécessaire. Enfin, l'effet de l'intervention est supposé identique pour tous les clusters et identique pour chaque période.

3.5.2 Les nouvelles méthodes non paramétriques

Plus récemment, d'autres méthodes non-paramétriques ont été proposées par (Kennedy-Shaffer et al., 2020).

Parmi elles, la méthode à contrôle synthétique est similaire à la méthode intra-période non paramétrique proposée par Thompson et ses collègues (Thompson et al., 2018). Cette dernière construit des comparaisons verticales, puis utilise une moyenne pondérée de ces comparaisons comme estimation de l'effet global.

Egalement, la méthode crossover, quant à elle, incorpore des comparaisons verticales au moment du crossover *(Figure 7).* Précisément, cette méthode compare le contraste moyen (ou effet moyen) entre la dernière période de contrôle et la première période d'intervention de chaque cluster au contraste moyenne des clusters restants dans le contrôle à ces mêmes périodes. Comme cette méthode inclue en plus une composante horizontale, cela permet de rétablir une part de la puissance perdue dans les approches exclusivement verticales.

Enfin, la méthode crossover-synthétique contrôle est une combinaison de cette dernière avec la méthode à contrôle synthétique.

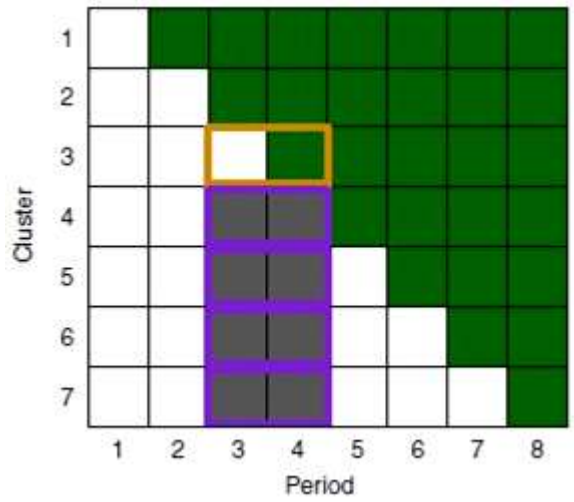


Figure 7 : Représentation de la méthode non-paramétrique « méthode crossover » (Kennedy-Shaffer et al., 2020).

4 L'ETUDE OREANE

4.1 Présentation de l'étude

4.1.1 Objectif principal

L'étude OREANE est un projet retenu à l'appel d'offre du PREPS pris en charge par l'UMES en collaboration avec l'USMR. L'objectif principal de l'étude est d'évaluer l'impact d'une formation complexe associée à des outils sur le taux d'erreur médicamenteuse des infirmières diplômées d'état (IDE) liées à l'administration de médicaments dans les Réanimations Néonatales. Au moment de la conception de l'étude, les connaissances dans les études stepped wedge étaient limitées, c'est pourquoi le schéma initialement prévu était de type « avant-après intervention». OREANE a finalement été convertie en un stepped wedge avec périodes de transition en 2015, avant la mise en place de l'étude dans le but de permettre à tous les centres de recevoir l'intervention. Le niveau du cluster considéré est le Centre Hospitalier Universitaire et les CHU de trois villes ont été inclus (Nîmes, Bordeaux et Caen). Il s'agit de la première étude prise en charge à l'UMES-USMR suivant ce type de schéma.

4.1.2 Les erreurs médicamenteuses

Les erreurs médicamenteuses sont le type d'erreurs le plus commun affectant la sécurité des patients et sont la cause la plus fréquente d'évènements indésirables évitables. Elles sont actuellement largement sous-déclarées. Pourtant, elles peuvent avoir des conséquences importantes, d'une part, pour la sécurité du patient, et d'autre part, pour la société (coûts, procédures judiciaires, etc.). La prescription et l'administration sont les deux étapes qui génèrent le plus d'erreurs médicamenteuses. Les erreurs d'administration sont bien connues et la survenue de la quasi-totalité de celles-ci est associée à des facteurs humains et organisationnels. Les taux d'erreurs médicamenteuses en pédiatrie et néonatalogie rapportés dans les études précédentes sont assez variables en raison de différences méthodologiques. De plus, la fréquence des erreurs dans cette population à risque est plus forte (jusqu'à trois fois plus) en regard des patients adultes. Ce fort risque d'erreurs découle :

- d'un environnement de travail critique ;
- de l'utilisation prépondérante (60%) de la voie intraveineuse ;
- d'une très grande fréquence d'utilisation des médicaments hors indications AMM (94% des médicaments prescrits en Unité de Soins Intensifs Néonatale : USIN) ;
- des particularités physiologiques des patients qui constituent une population très vulnérable.

L'erreur est définie comme un écart par rapport à ce qui aurait dû être fait au cours de la prise en charge thérapeutique médicamenteuse du patient. La mesure du taux d'erreur a été faite par vérification de type check-list (audit) par une personne extérieure à l'unité de soins (IDE de Recherche Clinique ou IDERC). Le taux d'erreurs de préparation et d'administration (π) est défini ainsi :

$$\pi = \frac{n_{err}}{n_{actes}} * 100$$

avec n_{err} le nombre d'erreurs détectées et n_{actes} le nombre total d'opportunités d'erreurs.

4.1.3 Définition de l'intervention

L'intervention mise en place dans l'étude OREANE, complexe, multifactorielle et progressive comporte une phase intensive suivie de phases d'entretien.

La phase intensive associe une formation en ligne (documents sous forme de Power Point animés, films, quizz, etc.) associée à des ateliers en présentiel d'une demi-heure ayant lieu une fois par semaine pendant 9 semaines. L'accès en ligne à une plateforme par les IDE incluses a été possible à tout moment durant la période de formation grâce à un code personnel. Des échanges sur les notions théoriques abordées lors des formations en ligne, une mise à disposition d'outils pratiques ainsi qu'un partage de méthodes de travail ont été permis lors des ateliers en présentiel.

Les phases d'entretiens ont fait appel à des formations pratiques sous forme d'ateliers en présentiel sur une durée d'une demi-heure, au rythme de deux sessions par mois (soit un total de 4 à 5 sessions sur les 9 semaines). Durant celles-ci, les documents de formations en e-learning ont été laissés à disposition des IDE.

La possibilité d'adaptation locale des outils de formations est un élément majeur mis en avant dans l'étude OREANE. En effet, les présentations et contenus n'ont pas été imposés car pouvant être perçus comme non adaptés par les infirmières au contexte local. De même, la fréquence des ateliers lors des phases d'entretien était suggérée, mais non imposée. La flexibilité du contenu de l'intervention avait pour objectif de faciliter son acceptabilité par les IDE. La participation active des infirmières aux ateliers est un autre élément majeur de l'étude (approche participative).

4.1.4 Schéma d'étude

Chaque intervention intensive et d'entretien a été introduite sur une période de neuf semaines dans chacun des centres de manière décalée durant les périodes de transition. Les interventions d'entretien ont été prévues à la suite de l'intervention intensive dans le but de

maintenir voire d'accentuer l'effet de la formation sur la fréquence des erreurs médicamenteuses, mais également pour permettre aux IDE absentes aux ateliers précédents de bénéficier d'une formation. La longue durée de l'introduction des interventions a permis une acquisition progressive des connaissances, et était nécessaire, par ailleurs, afin de former le maximum d'infirmières sans perturber le fonctionnement des activités de soins. Quant aux périodes d'observation, chacune d'entre elles a duré sept semaines. La première période s'est déroulée du 20 Novembre 2017 au 07 Janvier 2018, la seconde du 12 Mars au 29 Avril 2018, la troisième du 02 Juillet au 19 Aout 2018, la quatrième du 22 Octobre au 09 Décembre 2018 et la cinquième du 28 Janvier au 17 Mars 2019. La figure 8, ci-dessous, illustre le schéma de l'étude.

Centre	**OBS-1** (7 Sem)	**IMP** (9 Sem)	**OBS-2** (7 Sem)	**IMP** (9 Sem)	**OBS-3** (7 Sem)	**IMP** (9 Sem)	**OBS-4** (7 Sem)	**IMP** (9 Sem)	**OBS-5** (7 Sem)	**Total**
CHU de Nîmes	N_{ide}=24 N_{obs}=367	**Int**	N_{ide}=23 N_{obs}=259	Ent 1	N_{ide}=21 N_{obs}=331	Ent 2	N_{ide}=21 N_{obs}=359	Ent 3	N_{ide}=21 N_{obs}=333	N_{ide}=30 N_{obs}=1649
CHU de Bordeaux	N_{ide}=33 N_{obs}=298		N_{ide}=28 N_{obs}=368	**Int**	N_{ide}=23 N_{obs}=361	Ent 1	N_{ide}=18 N_{obs}=337	Ent 2	N_{ide}=24 N_{obs}=451	N_{ide}=41 N_{obs}=1815
CHU de Caen	N_{ide}=30 N_{obs}=348		N_{ide}=29 N_{obs}=369		N_{ide}=17 N_{obs}=241	**Int**	N_{ide}=28 N_{obs}=374	Ent 1	N_{ide}=25 N_{obs}=354	N_{ide}=53 N_{obs}=1686

Figure 8 : Schéma de l'étude OREANE. N_{ide} : Nombre d'IDE différentes, N_{obs} : Nombre d'observations différentes correspondant aux actes des IDE, Int : Intervention intensive, Ent1, 2, 3 : Interventions d'entretien, OBS : Phases d'observations, IMP : Phases d'implémentations des interventions.

La possibilité pour l'IDE de faire une erreur médicamenteuse a été recueillie à chaque observation correspondant à un acte infirmier. Chacune des IDE a été observée plusieurs fois au sein d'une même période et à des périodes différentes. D'autre part, des IDE ont pu arriver en cours d'étude et d'autres partir. Le schéma d'OREANE est donc de type cohorte ouverte car les données sont répétées au niveau individuel (IDE).

4.1.5 Particularités de l'étude OREANE

L'étude présente certaines limites :

- OREANE ne contient que trois clusters, ce qui est considéré comme insuffisant pour un schéma stepped wedge. Cependant, il y a grand nombre d'observations par cluster et par période (environ 343 en moyenne et 24 IDE en moyenne).
- L'intervention est complexe et dépendante de la durée d'exposition. En effet, les interventions d'entretien se sont produites à distance de l'intervention principale.

D'autre part, elle n'a pas été homogène dans tous les centres car les CHU de Nîmes, de Bordeaux et de Caen n'ont pas reçu le même nombre d'entretiens.

- Il existe quatre périodes de transition pour chaque cluster. De plus, les périodes de transition n'ont pas la même durée que les périodes d'observation.
- Le temps exact de réalisation des actes médicamenteux par les IDE n'a pas été recueilli. En effet, le seul temps à disposition dans la base de données est la période durant laquelle l'IDE est observée, ce qui est une information résumée car chaque période d'observation a une durée de sept semaines. Cela ne permet donc pas l'estimation de l'effet précis du temps avant et après formations sur la probabilité d'erreurs médicamenteuses des IDE.

4.2 Modélisation statistique de l'étude OREANE

La modélisation de l'impact de l'intervention sur les taux d'erreurs médicamenteuses des IDE liées à l'administration dans les Réanimations Néonatales a nécessité un ajustement sur une tendance temporelle pour éviter un éventuel phénomène de confusion. Par ailleurs, la répétition des données au niveau du cluster et au niveau individuel a impliqué une prise en compte de ces deux niveaux de corrélations. La stratégie de modélisation de l'étude OREANE a été décomposée en deux étapes. La première a consisté à trouver le modèle sans covariable le plus adapté aux données de l'étude. Une fois cette étape réalisée, les covariables d'ajustement ont été introduites dans le but d'obtenir le modèle final.

4.2.1 Les variables du modèle sans covariable

La variable à expliquer de l'étude OREANE Y_{ijk} est binaire : erreur médicamenteuse (Oui/Non) pour l'individu k du cluster i à l'observation j. Les observations j portent sur les actes infirmiers réalisés sur des enfants de moins de deux ans par des IDE ayant accepté de participer à l'étude. Les IDE peuvent être observées plusieurs fois au sein d'une même période, ainsi qu'à des périodes différentes.

La tendance temporelle a été introduite dans le modèle sous la forme de quatre variables indicatrices correspondant à une période d'observation donnée : $T1_{ijk}$, $T2_{ijk}$, $T3_{ijk}$ et $T4_{ijk}$. La classe de référence correspond à la période où tous les clusters sont dans la condition contrôle.

Concernant l'effet de l'intervention, les différentes modalités de l'intervention (intensive et entretiens) n'ont pas été prises en compte pour des problèmes d'identifiabilité des paramètres. En effet, les interventions d'entretiens entrent en compétition avec les différentes périodes d'observation. La colinéarité alors existante entre ces paramètres a impliqué de faire

un choix entre la présence d'un effet période d'une part et les interventions d'entretien d'autre part. En accord avec l'équipe investigatrice, le choix s'est arrêté sur la présence d'un effet période, d'une part pour éviter un phénomène de confusion, mais également car les trois centres n'ont pas bénéficié du même nombre d'interventions d'entretien. Il s'agit donc de l'estimation d'un effet global, ce qui constitue un résumé de l'information sous la forme d'une variable indicatrice Int_{ijk}=1 si le cluster est dans la condition « intervention » et 0 s'il est dans la condition « contrôle ».

Enfin, le centre de référence est celui de Nîmes, et les variables $C1_i$ et $C2_i$ sont respectivement les variables indicatrices pour le CHU de Bordeaux et celui de Caen.

4.2.2 Modélisation sans covariable de l'étude OREANE

Le modèle sans covariable choisit pour modéliser l'effet de l'intervention sur le taux d'erreur médicamenteuse des IDE est le suivant :

$$Logit[Pr(Y_{ijk}=1)]= \beta_0 + \alpha_{0ik} + \beta_1 C1_i + \beta_2 C2_i + \beta_{3à6} T1à4_{ijk} + \beta_7 Int_{ijk} + \beta_8 Int_{ijk}*C1_i + \beta_9 Int_{ijk}*C2_i$$

Avec α_{0ik}, l'intercept aléatoire du sujet k inclus dans le cluster i tel que $\alpha_{0ik} \sim N(0, \sigma_0^2)$.

Il s'agit d'un modèle logistique mixte car la variable « erreur médicamenteuse » est binaire. L'intercept aléatoire α_{0ik} permet de prendre en compte la corrélation des données répétées au niveau de l'IDE. L'ajout d'effets aléatoires supplémentaires à ce niveau tels qu'un effet temps aléatoire ou un effet traitement aléatoire ne permettait pas au modèle de converger. La prise en compte de la corrélation des données répétées au niveau du cluster a été introduite dans les effets fixes. En effet, les effets aléatoires supposent une distribution normale, ce qui est une hypothèse trop forte étant donné le faible nombre de clusters dans l'étude. Les termes associés aux coefficients β_8 et β_9 représentent les interactions entre l'intervention et le centre considéré. Le descriptif des données a montré un effet centre important, notamment dans la réponse à l'intervention, c'est pourquoi les termes d'interaction ont été introduits dans le modèle *(voir partie 4.3.1.4)*. L'effet des périodes d'observation associées aux termes allant de β_3 à β_6 ont permis d'ajuster l'effet de l'intervention sur un potentiel effet de confusion du temps (Hemming et al., 2015a).

Dans l'idéal, il aurait pu être intéressant d'ajouter les différentes interventions d'entretien au modèle (entretien 1, 2 et 3), ce qui aurait pu être envisageable si le temps avait été recueilli de manière continue. Une autre stratégie envisageable aurait été de modéliser l'interaction de l'effet de l'intervention avec le temps étant donné que celle-ci s'inscrit dans une démarche progressive. En effet, il est possible que l'effet de l'intervention se soit intensifié avec le

temps grâce à l'effet des interventions d'entretien. Malheureusement, une telle interaction n'a pas pu être estimée pour des problèmes de convergences.

D'autre part, il pourrait être imaginé que la tendance temporelle dépende du centre, ce qui aurait pu être pris en compte en introduisant des termes d'interaction entre ces derniers et les périodes d'observations. Cependant, ceci entraîne une déficience de rang dans les effets fixes (plus de paramètres à estimer que de possibilités différentes dans la base de données).

4.2.3 Modélisation avec covariables de l'étude OREANE

Les covariables d'ajustement ont été ajoutées au modèle initial pour obtenir les résultats du modèle final. Les facteurs d'ajustement retenus par la première sélection clinique sont : le type d'administration, l'ancienneté de l'IDE dans la profession, l'ancienneté de l'IDE dans le service, l'anticipation de la préparation du traitement, l'interruption de la tâche préparation lors de la préparation/administration ainsi que l'erreur de prescription. Toutes ces covariables ont été recueillies de manière qualitative et ont par conséquent été introduites dans la modélisation grâce à des variables indicatrices.

4.3 Résultats l'étude OREANE

4.3.1 Descriptif des données

4.3.1.1 Descriptif des établissements

Tableau 1 : Participation des IDE aux différentes formes de l'intervention selon le centre.

	CHU de Nîmes	CHU de Bordeaux	CHU de Caen
Taux de participation à l'intervention intensive	37,0%	24,0%	9,0%
Taux de participation à l'intervention d'entretien 1	31,0%	25,0%	Non fourni
Taux de participation à l'intervention d'entretien 2	37,0%	32,0%	-
Taux de participation à l'intervention d'entretien 3	40,0%	-	-
Participation à au moins 1 atelier des IDE incluses	93,3%	61,9%	22,2%
Au moins 1 QCM posté (formation en ligne) pour IDE incluses	71,4%	65,8%	28,3%

Comme observé dans le tableau ci-dessus, les taux de participations aux différentes interventions (intensives et entretiens) sont variables selon les centres. Dans l'hypothèse où l'intervention a un impact sur le taux d'erreurs médicamenteuses, un effet différent selon le

centre considéré pourrait alors être observé. En effet, les taux de participation aux différentes phases de l'intervention sont moins importants à Bordeaux qu'à Nîmes, et le taux de participation à l'intervention intensive du CHU de Caen est beaucoup plus faible par rapport aux deux autres centres (9,0%). Toutefois, ces faibles taux ne nuisent pas à l'évaluation de l'effet de l'intervention car l'étude OREANE s'intéresse à l'impact de la formation complexe « dans la vraie vie ». La variabilité des taux de participations au sein d'un même centre, eux, peuvent dépendre de facteurs saisonniers (moins de personnel présent pendant les vacances, activité accrue pendant certaines périodes, etc.).

Le taux de participation à au moins un atelier des IDE incluses ainsi que le taux d'IDE ayant posté au moins 1 QCM sont dépendants de la durée de l'exposition. En effet, il est logique que le taux de participation des IDE à au moins un atelier ou un QCM soit plus important au CHU de Nîmes (trois phases d'entretien réalisées) par rapport au CHU de Bordeaux (deux phases d'entretiens réalisées), lui-même plus important que le CHU de Caen (une seule phase d'entretien réalisée).

4.3.1.2 Descriptif des actes médicaux réalisés par les IDE

Comme illustré grâce au descriptif ci-dessous, la grande majorité des actes médicaux réalisés par les IDE dans le cadre de l'étude OREANE correspondent à des traitements administrés par voie parentérale (78,80%). D'autre part, les actes des IDE ayant une ancienneté professionnelle d'au moins six ans représentent plus de 65% des actes. En moyenne, presque 50% des actes ont été réalisés par des IDE ayant au moins six ans d'ancienneté dans leur service actuel. Les deux covariables relatives à l'ancienneté des IDE sont les seules qui comportent des données manquantes (pour 442 observations chacunes). En outre, seulement 9,36% des traitements ont été préparés de manière anticipée. Durant l'étude, 34,54% des tâches correspondant aux actes ont été interrompues, ce qui reflète le climat d'urgence dans ces unités. Enfin, plus d'un quart des observations ont présenté une erreur dans la prescription (incomplète ou illisible).

Tableau 2 : Descriptifs des covariables d'ajustement dans le cadre de l'analyse de l'étude OREANE.

Variable		Total		Pas d'erreur médicamenteuse		Erreur médicamenteuse	
Type d'administration							
	n (dm)	5150 (0)		1021 (0)		3804 (0)	
	Parentérale	4058	78,80%	449	43,98%	3309	86,99%
	Orale	1092	21,20%	572	56,02%	495	13,01%
Ancienneté dans la profession (IDE)							
	n (dm)	4708 (442)		957 (64)		3429 (375)	
	Moins de un an	258	5,48%	55	5,75%	194	5,66%
	Un à deux ans	302	6,41%	108	11,29%	180	5,25%
	Trois à cinq ans	795	16,89%	146	15,26%	584	17,03%
	Six à dix ans	1490	31,65%	331	32,50%	1087	31,70%
	Onze ans et plus	1863	39,57%	337	35,21%	1384	40,36%
Ancienneté dans le service actuel							
	n (dm)	4708 (442)		957 (64)		3429 (375)	
	Moins de un an	450	9,56%	104	10,87%	329	9,59%
	Un à deux ans	825	17,52%	178	18,60%	610	17,79%
	Trois à cinq ans	986	20,94%	222	23,20%	671	19,57%
	Six à dix ans	1141	24,24%	215	22,47%	860	25,08%
	Onze ans et plus	1306	27,74%	238	24,87%	959	27,97%
Préparation anticipée du traitement							
	n (dm)	5150 (0)		1021 (0)		3804 (0)	
	Non	4626	89,93%	972	95,20%	3406	89,54%
	Oui	482	9,36%	46	4,51%	370	9,73%
	Information non disponible	42	0,82%	3	0,29%	28	0,74%
Interruption de la tâche							
	n (dm)	5150 (0)		1021 (0)		3804 (0)	
	Non	3371	65,46%	665	65,13%	2464	64,77%
	Oui	1779	34,54%	356	34,87%	1340	35,23%
Erreur de prescription							
	n (dm)	5150 (0)		1021 (0)		3804 (0)	
	Non	3817	74,12%	808	79,14%	2749	72,24%
	Oui (incomplète ou illisible)	1333	25,88%	213	20,86%	1055	27,73%

4.3.1.3 Descriptif de la variable réponse

Comme le montre le descriptif ci-dessous en Figure 9, le pourcentage d'erreur médicamenteuse des IDE (Erreur médicamenteuse = oui) diminue à chaque période d'observation successive. Ceci semble aller dans le sens d'une efficacité de l'effet de l'intervention car plus la période de l'étude est avancée, plus le nombre de centre ayant reçu

l'intervention est important (dû à l'implémentation décalée inhérente au schéma stepped wedge).

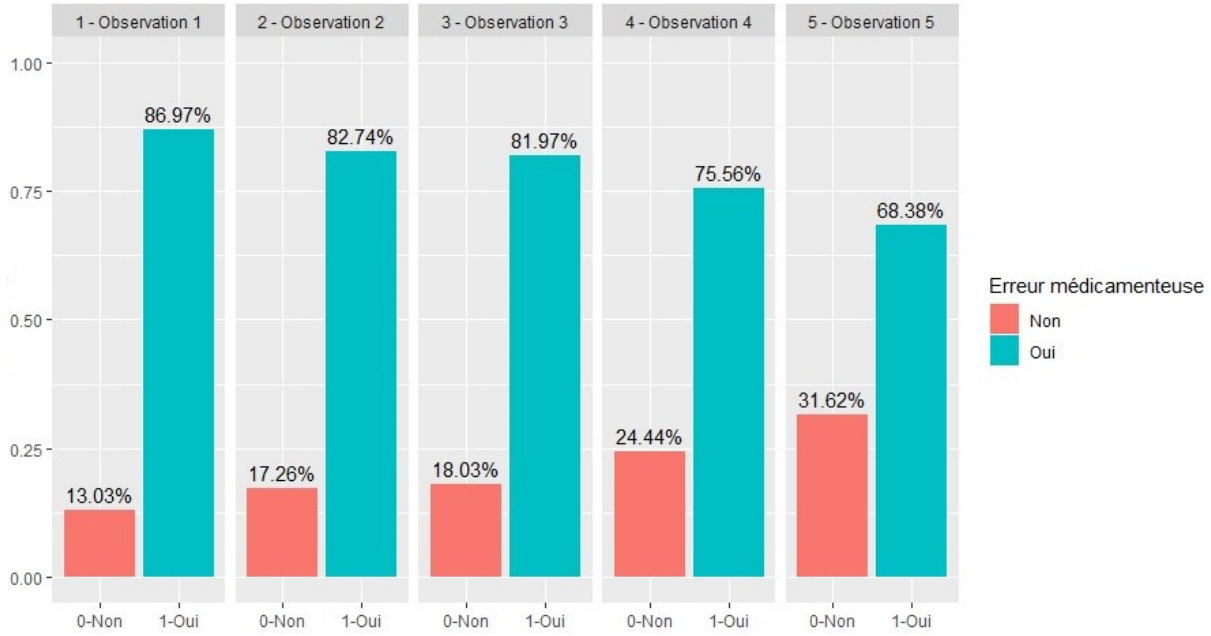


Figure 9 : Descriptif du taux d'erreur médicamenteuse des IDE par période d'observation.

4.3.1.4 Descriptif de la variable réponse par centre

Comme le présente la figure 10 ci-dessous, les taux d'erreurs avant l'intervention sont différents suivant le centre, ce qui montre une forte hétérogénéité. En effet, le taux d'erreur avant intervention au CHU de Bordeaux est en moyenne de 72,00%, ce qui est inférieur au CHU de Nîmes 96,20% et au CHU de Caen 99,00%. D'autre part, pour le CHU de Bordeaux, une augmentation du taux d'erreur avant l'intervention (de 62,10% à 79,90%) est observée d'après le graphique ci-dessous. Cette augmentation avant l'intervention ne peut pas se présenter au CHU de Nîmes car il reçoit l'intervention dès la seconde période. Elle n'est pas observée non plus au CHU de Caen qui a des taux d'erreurs initiaux très élevés ne pouvant donc pas augmenter de manière similaire. Les évolutions observées des taux d'erreur médicamenteuse avant l'intervention sont donc hétérogènes pour le CHU de Bordeaux et celui de Caen.

Le taux d'erreur au CHU de Nîmes après l'intervention se traduit par une diminution par rapport au taux avant l'intervention (de 96,20% avant à 63,10% à la période suivante). Par la suite, ce taux fluctue entre 69,00% et 55,00%, ce qui ne permet pas de dégager une tendance quelconque dans cette diminution. A priori, il n'y aurait pas d'intensification de l'effet de l'intervention intensive par les phases d'entretien dans ce centre, ce qui serait plus favorable à un effet immédiat.

La variation du taux d'erreur avant et après l'intervention au CHU de Bordeaux n'est pas identique. En effet, ce dernier ne diminue pas en moyenne à la suite de l'implémentation des interventions (72,00% avant et 74,10% après).

Enfin, comme au CHU de Nîmes, une diminution du taux d'erreur médicamenteuse est observée en moyenne après l'intervention (99,00% avant et 85,00% après). Cependant, la diminution la plus importante ne se produit pas directement après l'implémentation de l'intervention intensive, mais à la suite de l'intervention d'entretien 1 (94,34% à 75,73%).

En conclusion, une forte hétérogénéité des centres est observée, d'une part dans les taux avant l'intervention, et d'autre part dans la réponse à cette dernière.

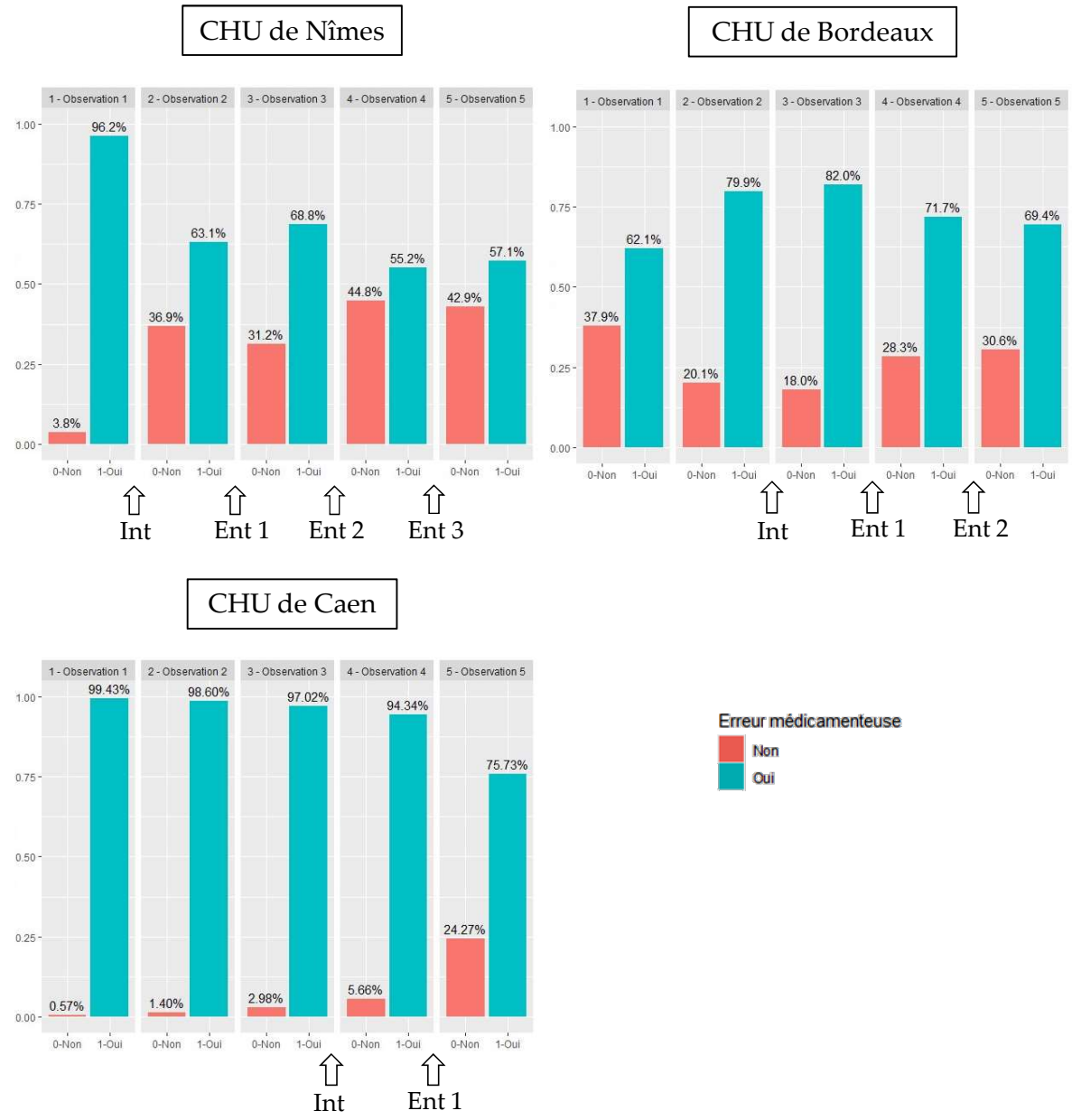


Figure 10 : Descriptif du taux d'erreur médicamenteuse des IDE par période d'observation par centre.

4.3.1.5 Les données manquantes de la variable réponse

Le tableau ci-dessous décrit les données manquantes concernant la variable « erreur médicamenteuse » par centre et par période d'observation. Au total, il y a 325 observations pour lesquelles la donnée est manquante, ce qui représente plus de 6% des actes (325/5150).

Tableau 3 : Représentation des données manquantes pour la variable « erreur médicamenteuse » par centre et par période d'observation.

Centre / Période	CHU de Nîmes	CHU de Bordeaux	CHU de Caen	Total
1	49 (15%)	5 (2%)	0 (0%)	54 (17%)
2	23 (7%)	0 (0%)	11 (3%)	34 (10%)
3	62 (19%)	0 (0%)	6 (2%)	68 (21%)
4	80 (25%)	1 (<1%)	3 (1%)	84 (26%)
5	67 (20%)	6 (2%)	12 (4%)	85 (26%)
Total	281 (86%)	12 (4%)	32 (10%)	325 (100%)

La grande majorité des données manquantes concernant la variable réponse sont retrouvées au CHU de Nîmes (86% des données manquantes). Il y a peu de données manquantes pour le CHU de Bordeaux et pour le CHU de Caen qui totalisent à eux deux 14% des données manquantes pour cette variable.

Par ailleurs, le pourcentage de données manquantes augmente avec la période (17% à la première et 26% à la dernière).

4.3.2 Résultats de la modélisation

4.3.2.1 Résultats du modèle sans covariable

Tous les rapports de côtes présentés ci-dessous dans les tableaux 4 et 5 s'interprètent comme « toutes choses étant égales par ailleurs » (après ajustement sur les autres variables du modèle).

Comme attendu, l'effet de l'intervention n'est pas le même en fonction du centre, même s'il est globalement significatif. Pour le CHU de Nîmes, l'effet de l'intervention est significatif car l'intervalle de confiance du rapport de côte ne comprend pas la valeur 1. Dans ce centre, l'exposition l'intervention a permis de diminuer le taux d'erreur médicamenteuse des IDE car la mesure d'association est inférieure à 1 (0,03). En revanche, ces résultats ne montrent pas d'efficacité de l'intervention pour le CHU de Bordeaux car l'intervalle de confiance du rapport de côte comprend la valeur 1, ce qui est en adéquation avec les représentations de taux d'erreur par centre. Enfin, l'effet de l'intervention est également significatif dans le sens d'une diminution du taux d'erreur pour les exposés au CHU de Caen, même si la mesure de

l'association est moins forte que pour le CHU de Nîmes. Toutefois, ce résultat n'est pas surprenant car les IDE du CHU de Caen ont été exposées moins longtemps que celles du CHU de Nîmes et le centre n'a bénéficié que d'une seule intervention d'entretien alors que le CHU de Nîmes en a bénéficié de trois. Enfin, la p-value pour la variance de l'intercept aléatoire est très significative, ce qui indique l'importance de la prise en compte des données répétées au niveau de l'IDE.

Tableau 4 : Présentation du modèle initial d'OREANE : étude de l'effet de l'intervention globale sur le taux d'erreur médicamenteuse des IDE dans les unités de réanimation néonatale.

AIC = 4219.7						
Strates	**Variables**	**Modalités**	**Effectif**	$\widehat{OR}$	**IC** 95%	**p-value**
	Phase d'observation					<0,0001
		OBS-1	959			
		OBS-2	962	2,00	[1,40 ; 2,86]	
		OBS-3	865	2,17	[1,33 ; 3,55]	
		OBS-4	986	1,50	[0,92 ; 2,45]	
		OBS-5	1053	1,08	[0,66 ; 1,76]	
CHU de Nîmes	Effet de l'intervention					<0,0001
		Non	318			
		Oui	1050	0,03	[0,01; 0,07]	
CHU de Bordeaux	Effet de l'intervention					
		Non	1142			
		Oui	661	0,91	[0,59 ; 1,41]	
CHU de Caen	Effet de l'intervention					
		Non	941			
		Oui	713	0,11	[0,05 ; 0,2]	

Effets aléatoires					
Variable	**Niveau hiérarchique**	**Effectif**	**Variance**	**Ecart-type de la variance**	**p-value**
Intercept	IDE	4825	0,33	0,08	<0,0001

La p-value pour l'effet global de l'intervention a été obtenue grâce au test du rapport de vraisemblance. La p-value pour la variance de l'intercept aléatoire a été obtenue grâce à un test de mélange de Chi-deux.

4.3.2.2 Résultats du modèle avec covariables

Toutes les variables sélectionnées cliniquement ont été insérées comme facteurs d'ajustement. Un modèle complet a été utilisé. Ainsi, les variables ne présentant pas d'association significative avec le taux d'erreur médicamenteuse n'ont pas été retirées du modèle.

Tableau 5 : Présentation du modèle final d'OREANE : étude de l'effet de l'intervention globale sur le taux d'erreur médicamenteuse des IDE dans les unités de réanimation néonatale avec covariables d'ajustement.

AIC = 3235.9						
Strates	**Variables**	**Modalités**	**Effectif**	$\widehat{OR}$	**IC** 95%	**p-value**
	Phase d'observation					<0,0001
		OBS-1	959			
		OBS-2	962	1,83	[1,20 ; 2,80]	
		OBS-3	865	0,81	[0,46 ; 1,42]	
		OBS-4	986	0,66	[0,37 ; 1,18]	
		OBS-5	1053	0,39	[0,22 ; 0,68]	
CHU de Nîmes	Effet de l'intervention					<0,0001
		Non	318			
		Oui	1050	0,05	[0,02; 0,11]	
CHU de Bordeaux	Effet de l'intervention					
		Non	1142			
		Oui	661	2,23	[1,30 ; 3,82]	
CHU de Caen	Effet de l'intervention					
		Non	941			
		Oui	713	0,42	[0,20 ; 0,87]	
	Type d'administration					<0,0001
		Parentérale	4058			
		Orale	1092	0,08	[0,06 ; 0,10]	
	Ancienneté dans la profession (IDE)					<0,001
		Moins de un an	258			
		Un à deux ans	302	0,31	[0,14 ; 0,73]	
		Trois à cinq ans	795	1,11	[0,44 ; 2,83]	
		Six à dix ans	1490	1,03	[0,42 ; 2,55]	
		Onze ans et plus	1863	0,86	[0,31 ; 2,41]	
	Ancienneté dans le service actuel					0,53
		Moins de un an	450			
		Un à deux ans	825	1.04	[0,51 ; 2,14]	
		Trois à cinq ans	986	0.74	[0,34 ; 1,61]	
		Six à dix ans	1141	0.87	[0,39 ; 1,98]	
		Onze ans et plus	1306	1.13	[0,44 ; 2,9]	
	Préparation anticipée du traitement					0,27
		Non	4626			
		Oui	482	0,84	[0,56 ; 1,26]	
		Information non disponible	42	2,38	[0,65 ; 8,66]	
	Interruption de la tâche					0,95
		Non	3371			
		Oui	1779	1	[0,81 ; 1,25]	
	Erreur de prescription					0,14
		Non	3817			
		Oui (incomplète ou illisible)	1333	0,84	[0,67 ; 1,06]	
Effets aléatoires						
Variable	**Niveau hiérarchique**	**Variance**	**Ecart-type de la variance**			**p-value**
Intercept	IDE	0,55	0,13			<0,0001
La p-value pour l'effet global de l'intervention a été obtenue grâce au test du rapport de vraisemblance. La p-value pour la variance de l'intercept aléatoire a été obtenue grâce à un test de mélange de Chi-deux.						

Les variables d'ajustement introduites ont fait diminuer de manière importante le critère d'Aïkake (AIC de 4219,7 à 3235,9), ce qui témoigne d'une bonne amélioration dans l'adéquation aux données.

En comparant les deux modèles, plusieurs changements importants dans les résultats peuvent être relevés. Concernant la variable « phase d'observation », l'effet lié à la troisième période par rapport à la période de référence (contrôle) a disparu car l'intervalle du rapport de côtes comprend la valeur 1 dans le modèle ajusté. En revanche, le résultat devient significatif pour la cinquième période d'observation par rapport à la période de contrôle dans le sens d'une diminution du taux d'erreur médicamenteuse, qui n'était pas observée dans le modèle sans covariable. Il n'y a toujours pas de tendance significative pour la quatrième période et le rapport de côte associé à la seconde est encore significativement supérieur à 1.

En ce qui concerne l'effet de l'intervention, celui-ci reste significatif pour les CHU de Nîmes et de Caen et vont dans le sens d'une diminution du taux d'erreur médicamenteuse après l'intervention, même si les résultats sont moins prononcés que ceux obtenus avec le modèle précédent, surtout pour le centre de Caen (Rapport de Côte de 0,11 à 0,42). De manière plus surprenante, le taux d'erreur médicamenteuse augmente de manière significative à Bordeaux après l'intervention car le rapport de côte est positif (2,23) et exclut la valeur 1.

Le type d'administration est un facteur significativement associé à la probabilité pour une IDE de faire une erreur médicamenteuse. En effet, la côte de l'erreur est plus faible pour les administrations de type orales que celles de type parentérales (rapport de côte plus petit que 1).

Quant à l'ancienneté professionnelle en tant qu'IDE, on met en évidence une association significative avec le taux d'erreur médicamenteuse, car la p-value globale est largement inférieure à 5%. Toutefois, il peut tout de même être noté que seul le rapport de côtes comparant la catégorie « un à deux ans d'ancienneté » versus « moins d'un an » est significativement différent de 1, et va dans le sens d'une probabilité plus faible de faire une erreur pour les IDE ayant une à deux années d'ancienneté. En revanche, les rapports de côtes des catégories associées à plus d'ancienneté ne sont pas significativement différents de 1.

En outre, aucune association significative n'est mise en évidence entre l'ancienneté de l'IDE dans un service, la préparation anticipée du traitement, l'interruption de la tâche ainsi que l'erreur de prescription avec le taux d'erreur médicamenteuse chez les IDE car les p-value sont toutes supérieures à 5%.

4.3.3 Adéquation et discrimination

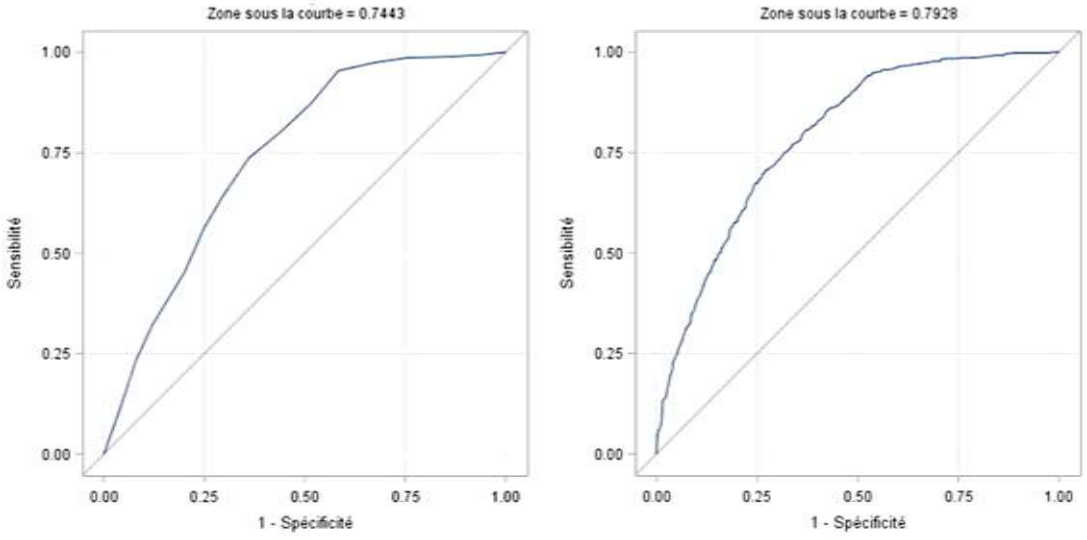
4.3.3.1 Adéquation des modèles

Les statistiques d'adéquation pour la distribution conditionnelle permettent d'évaluer la qualité d'ajustement sachant les effets aléatoires. Pour les deux modèles présentés, les ratios entre le Chi-deux de Pearson et les degrés de liberté sont de 0,99 pour le modèle sans covariable et de 1,04 pour le modèle avec covariables. Comme ces deux ratios sont très proches de 1, cela indique que la variabilité des données de l'étude a été correctement modélisée, et qu'il n'y a pas de surdispersion résiduelle.

4.3.3.2 Pouvoir discriminant des modèles

L'aire sous la courbe ROC (AUC) permet d'évaluer le pouvoir discriminant du modèle logistique. La courbe ROC peut être tracée en calculant les probabilités prédites par le modèle. Ces dernières peuvent être calculées grâce au « Best Linear Unbiaised Prediction » (BLUP), ou bien en se basant seulement sur les effets fixes. La figure 11 ci-dessous est une représentation de ces indicateurs pour les modèles sans covariable et avec covariables.

Pouvoir discriminant du modèle sans covariable



Prédictions marginales **Prédictions avec le BLUP**

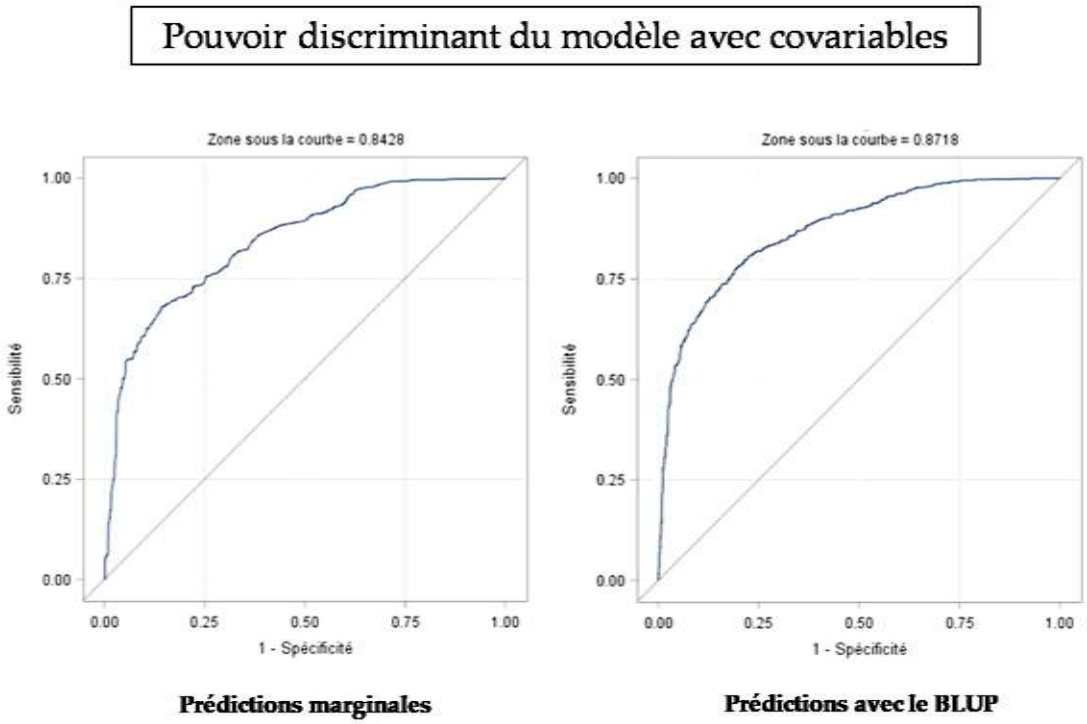


Figure 11 : Représentation des pouvoirs discriminants des modèles sans et avec covariables de l'étude OREANE.

Comme constaté d'après la figure ci-dessus, le pouvoir discriminant des modèles sans covariable et avec covariables est meilleur lorsque les effets aléatoires sont pris en compte pour les prédictions (BLUP). Par ailleurs, l'insertion des variables d'ajustement a également augmenté le pouvoir discriminant du modèle car il passe de 0,79 à 0,87, ce qui représente une excellente discrimination.

4.3.4 Application de l'analyse non paramétrique intra-période à l'étude OREANE

La méthode non paramétrique d'analyse intra-période a été appliquée aux données de l'étude OREANE à titre illustratif. Toutefois, il ne s'agit pas de la méthode retenue pour la communication des résultats finaux de l'étude pour des raisons de non adéquation de la méthode à l'étude OREANE et aux stepped wedge de type cohorte de manière plus générale. Les raisons précises seront vues dans la discussion des résultats.

Pour l'étude OREANE, cette méthode consiste à comparer les taux d'erreur médicamenteuse des conditions de contrôle et d'intervention, au sein des périodes contenant à la fois des clusters dans ces deux conditions d'exposition. Dans le cas ci-présent, il s'agit des périodes deux et trois *(Figure 12).* La figure ci-dessous représente les taux d'erreur médicamenteuse en fonction de la condition d'exposition pour les données regroupées de la période deux et trois.

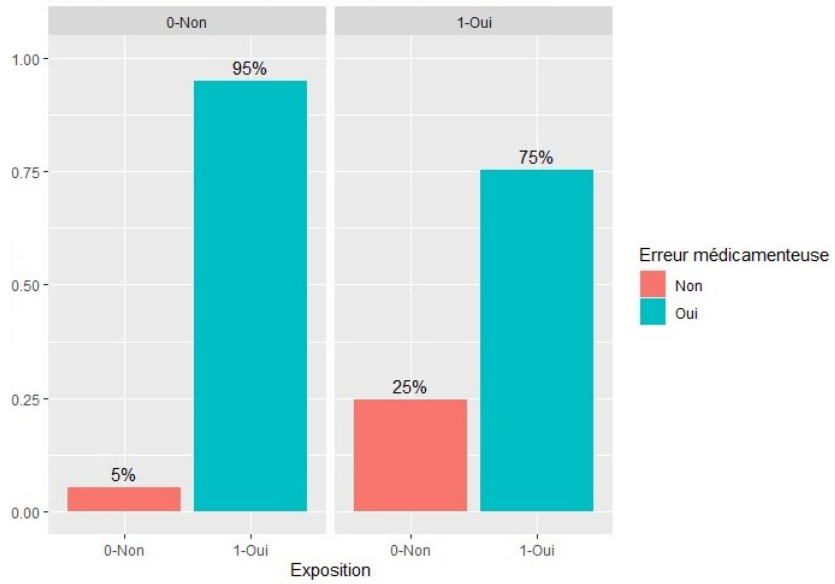


Figure 12 : Présentation des taux d'erreur médicamenteuse avant et après l'intervention pour les données regroupées de la période deux et trois.

La première étape de cette méthode consiste à résumer les informations de la variable réponse (erreur médicamenteuse : oui/non) pour chaque période et chaque condition vis-à-vis de l'intervention.

Tableau 6 : Résumé de la variable réponse de l'étude OREANE en fonction de la période et du statut vis-à-vis de l'intervention.

Période	Intervention	$\hat{\pi}$	$Logit\,(\hat{\pi})$	Effectif(n)	$Var[Logit\,(\hat{\pi})]$
1	Non	0,87	1,90	959	2,06
2	Non	0,89	2,10	726	2,23
3	Non	0,97	3,48	235	3,51
2	Oui	0,63	0,54	236	1,19
3	Oui	0,76	1,72	630	1,51
4	Oui	0,76	1,13	986	1,48
5	Oui	0,68	0,77	1053	1,29

$\hat{\pi}$ correspondant à la proportion moyenne d'erreur médicamenteuse dans chaque condition. L'effectif ici correspond aux nombres d'actes observés dans chacune des conditions. La variable réponse Y_{ijk} étant binaire, l'effet de l'intervention sera estimé dans l'échelle du logit. L'effet de l'intervention s'interprète comme une différence moyenne et pondérée entre la condition d'intervention et la condition contrôle.

Le tableau ci-dessous présente les différences estimées ainsi que leurs pondérations respectives :

Tableau 7 : Application de l'analyse intra-période aux données de l'étude OREANE.

Période	$g(\widehat{\pi}_0)$	n_0	$var[g(\widehat{\pi}_0)]$	$g(\widehat{\pi}_1)$	n_1	$var[g(\widehat{\pi}_1)]$	$g(\widehat{\pi}_1)$- $g(\widehat{\pi}_0)$= $\widehat{\theta}_j$	$\widehat{w}_j$
1	1,90	959	2,06	-	-	-	-	-
2	2,10	726	2,23	0,54	236	1,19	-1,56	90,20
3	3,48	235	3,51	1,17	630	1,51	-2,31	83,38
4	-	-	-	1,13	986	1,48	-	-
5	-	-	-	0,77	1053	1,29	-	-

avec g(x) comme fonction de lien, c'est-à-dire la fonction logit. $\widehat{\boldsymbol{\pi}}_{\mathbf{0}}$ correspond à la proportion moyenne d'erreur médicamenteuse dans la condition de contrôle et $\widehat{\boldsymbol{\pi}}_{\mathbf{1}}$ dans la condition d'intervention. $\widehat{\boldsymbol{w}}_{\boldsymbol{j}}$ correspond à la pondération associée à la période j.

L'estimation de l'effet de l'intervention correspond à la moyenne pondérée des différences estimées entre les conditions de contrôle et d'intervention :

$\hat{\theta}=\sum_{j=1}^{T}\frac{w_j}{w}\hat{\theta}_j$= 0,52*-1,52 + 0,48*-2,31 =-1,90

Cet effet de l'intervention étant estimé dans l'échelle du logit, on retrouve le rapport de côtes comparant les exposés aux non exposés par la formule $\widehat{\boldsymbol{OR}} = \exp^{(\widehat{\theta})} = 0{,}15$, un rapport de côtes étant l'exponentielle d'une différence de logit. En l'occurrence, il s'agit d'un rapport de côtes particulier car il dépend de pondérations.

Comme vu précédemment en utilisant les modèles linéaires généralisés, la mesure de l'association va dans le sens d'une diminution du taux d'erreur chez les exposés à l'intervention par rapport aux non exposés. Le test de permutation permet de savoir si ce rapport de côtes est significativement différent de 1.

Le principe de la permutation est de changer aléatoirement l'ordre des clusters et donc la période à laquelle ceux-ci passeront de la condition de contrôle à la condition d'intervention. La p-value est déterminée par la proportion d'effets de l'intervention après permutation étant au moins aussi important en valeur absolue que l'effet $\hat{\theta}$ calculé ci-dessus.

En réalisant dix-milles permutations, une p-value de 0,1669 est obtenue, ce qui ne permet pas de conclure à une association significative entre l'effet de l'intervention et le taux d'erreur médicamenteuse. En effet, le nombre de permutations possibles dans cet exemple est de 3! = 6 (3 correspondant au nombre de clusters). Ces possibilités sont représentées dans le tableau ci-dessous :

Tableau 8 : Possibilités de permutations dans l'étude OREANE avec trois clusters (3! = 6).

Possibilités	**1**	**2**	**3**	**4**	**5**	**6**
Ordre des centres	Nîmes Bordeaux Caen	Nîmes Caen Bordeaux	Bordeaux Nîmes Caen	Bordeaux Caen Nîmes	Caen Nîmes Bordeaux	Caen Bordeaux Nîmes

Par ailleurs, le ratio 1/6 est égal à 0,1666. D'après la p-value obtenue, il peut alors en être déduit qu'il y a au moins un effet estimé après permutation au moins aussi important que l'effet $\hat{\theta}$. Le grand nombre de permutations réalisées (10 000 le cas présent), permet de s'assurer que les six possibilités soient représentées de manière équitable grâce à la théorie des grands nombres.

5 DISCUSSION ET CONCLUSION

5.1 Discussion

5.1.1 Le schéma stepped wedge

Le schéma stepped wedge constitue un schéma innovant dans le cadre de l'évaluation d'une intervention ou d'un service au niveau du cluster comme en témoigne la littérature croissante dans ce domaine. Il confère différents avantages sur le plan statistique et méthodologique qui expliquent ce développement. Même s'il n'était encore que peu utilisé il y a quelques années, les projets adoptant ce type de schéma sont en plein essor. Principalement, il a été montré que ce schéma nécessite moins de clusters par rapport au schéma parallèle lorsque la corrélation ICC est grande et que la taille des clusters est importante (Barker et al., 2017). Mais ce schéma permet également d'implémenter l'intervention à tous les clusters, ce qui est un critère fondamental facilitant le recrutement et donc la réalisation d'un projet de recherche. Outre les aspects statistiques, le stepped wedge offre des solutions sur les plans logistique et éthique.

5.1.2 Analyse statistique des stepped wedge

Un grand nombre d'approches statistiques différentes (modèles mixtes, approche GEE, approches non paramétriques, etc…) est également retrouvé, ce qui illustre d'une part des diversités dans le schéma d'étude et d'autre part la volonté de répondre au mieux aux problèmes méthodologiques inhérents à ce dernier. Les biais des différentes méthodes statistiques sont des points souvent abordés (biais d'estimation, puissance, erreur de type I, etc…). Au même titre, d'autres aspects statistiques sont également détaillés comme le calcul du nombre de sujets nécessaires par exemple, où des applications en ligne ont été développées sur R (Hemming et al., 2020) et (Voldal et al., 2020).

Par ailleurs, les essais stepped wedge de type cohorte sont bien moins souvent retrouvés dans la littérature par rapport aux stepped wedge de type transversal. Il en résulte que les méthodes ainsi que les problèmes statistiques sont moins abondants pour ce schéma particulier.

Enfin, les approches non paramétriques récemment décrites dans la littérature (Kennedy-Shaffer et al., 2020), représentent des alternatives pour les études stepped wedge de type transversal contenant peu de clusters (six au minimum). Les approches utilisant une comparaison à la fois verticale et horizontale permettent d'améliorer la puissance de l'analyse par rapport aux approches avec une seule comparaison verticale. L'intérêt de ces dernières est de s'affranchir des hypothèses des méthodes paramétriques qui peuvent ne pas être vérifiées dans les études avec peu de clusters.

5.1.3 Les taux d'erreur médicamenteuse

Le descriptif des données concernant les proportions d'erreur médicamenteuse par centre dans les unités de soins de réanimation néonatale a révélé des taux d'erreur bien plus élevés que ceux attendus initialement (20%). En effet, les taux observés avant intervention sont de l'ordre de plus de 85% en moyenne, ce qui paraît considérable. Les discussions avec l'équipe investigatrice ont permis d'apporter quelques réponses sur ce point. La littérature dans ce domaine communique des résultats fortement hétérogènes, et des taux d'erreurs aussi importants ont déjà été retrouvés.

En premier lieu, de nombreux critères sont pris en compte dans la définition d'une erreur médicamenteuse. De plus, il suffit qu'un seul de ces critères soit défaillant pour considérer l'erreur médicamenteuse comme avérée. Cela constitue un résumé très strict de l'acte médical et ne le représente pas dans sa globalité, ce qui dresse un tableau sans doute plus négatif que la réalité. D'autre part, les possibilités de faire une erreur médicamenteuse ont été recueillies par une seule personne différente dans chaque centre. Il est alors probable qu'il existe une variabilité inter-observateur due principalement à des méthodes d'observations et/ou des sensibilités variables. Enfin, l'étude OREANE se base sur une approche participative fortement dépendante du centre considéré.

Les discussions avec l'équipe investigatrice ont également permis de souligner que les attitudes vis-à-vis de l'intervention à l'étude ont été sensiblement différentes entre les centres. Notamment, le centre de Nîmes est celui qui a montré le plus d'implication et de collaboration participative, ce que l'on retrouve dans la partie « taux de participation » du descriptif. En revanche, le centre de Bordeaux s'est montré beaucoup moins impliqué et a été victime d'une forte rotation du personnel. Ces derniers éléments peuvent sans doute expliquer en partie les résultats différents observés pour le CHU de Bordeaux concernant l'effet de l'intervention. En effet, la rotation du personnel peut nuire à un effet positif de l'intervention si les IDE formées ne restent pas dans le service. Il peut donc s'agir d'un facteur de confusion, car on peut penser qu'une personne nouvellement recrutée est en période d'adaptation et donc potentiellement plus à risque de faire une erreur. De plus, la rotation en tant que phénomène global et important à l'échelle d'une structure pourrait avoir eu des conséquences plus profondes (cohésion de l'équipe et/ou mécontentement au travail). Il est également possible que la probabilité de faire une erreur médicamenteuse soit liée à ce phénomène de rotation du personnel.

Concernant le CHU de Caen, l'équipe investigatrice a rapporté une mise en place de l'intervention plus lente par rapport aux autres centres, ce qui peut être observé sur le descriptif des données où l'on voit une réelle diminution du taux d'erreur qu'après la première phase d'entretien.

5.1.4 Le faible nombre de clusters dans l'étude OREANE

L'étude OREANE n'a pas été initialement conceptualisée en tant que stepped wedge. La première des conséquences de ce changement méthodologique est le faible nombre de clusters (trois). Ce faible nombre ne permet pas de conclure à un effet général de l'intervention à ce niveau. Le nombre de clusters minimum généralement considéré dans la littérature est de six.

Ce changement a également eu des impacts sur la modélisation statistique. En effet, comme décrit dans la partie « Modélisation statistique », les clusters sont généralement pris en compte par des effets aléatoires d'autant plus si leur nombre est élevé, ce qui implique des hypothèses de normalité asymptotique que l'on ne peut pas faire avec seulement trois clusters. Ceci explique pourquoi ils ont été modélisés dans les effets fixes pour l'étude OREANE. Si le nombre de clusters avait été suffisamment important, un modèle multi-niveaux de régression logistique mixte aurait probablement été considéré pour prendre en compte à la fois le niveau du cluster et le niveau individuel dans les effets aléatoires. En effet, il s'agit du type de modèle le plus retrouvé et décrit pour les stepped wedge de type cohorte.

5.1.5 Effet temps aléatoire et/ou effet traitement aléatoire

Différents choix d'effets aléatoires sont possibles et décrits dans la littérature à ce jour concernant l'analyse des stepped wedge. Notamment, un effet temps et/ou traitement aléatoire(s) constitue(nt) la(es) principale(s) possibilité(s). En pratique, il est recommandé d'inclure dans la mesure du possible ces deux effets aléatoires. Cette solution permet de mieux prendre en compte la corrélation des données sans pour autant avoir d'impact sur l'estimation des effets fixes. Toutefois, cette option n'est pas toujours réalisable pour des problèmes de convergence, ce qui implique de choisir l'un ou l'autre.

Dans le cas de l'étude OREANE, l'interaction entre l'effet de l'intervention et le centre peut être considérée comme l'alternative à un effet traitement aléatoire comme cela aurait été le cas si le centre avait été modélisé dans les effets aléatoires.

Toutefois, les discussions aujourd'hui ont plutôt tendance à recommander de modéliser un effet temps aléatoire qu'un effet traitement aléatoire dans le cas où il y a peu de séquences dans l'étude. L'étude OREANE ayant quatre séquences est dans ce cas de figure, cependant, un terme d'interaction entre le centre et la période ne pouvait pas être ajouté dans les effets fixes pour des problèmes de déficience de rang. En effet, comme le temps est considéré de manière catégorielle, l'interaction est représentée par l'ajout de huit paramètres dans le modèle (2*4).

De manière plus générale, la prise en compte d'un effet temps aléatoire peut également s'avérer difficile lorsque le temps est considéré de manière catégorielle, car cela signifie qu'il faut autant de paramètres dans les effets aléatoires qu'il y a de séquences dans l'étude, ce qui

peut poser des problèmes de convergence dus à une surparamétrisation. Pour OREANE, si le temps avait été recueilli de manière continue, nous aurions pu éventuellement le modéliser de manière quantitative après avoir étudié l'hypothèse de linéarité avec le logit. Il aurait été alors possible de modéliser une interaction entre le temps et le centre qui aurait mieux pris en compte l'hétérogénéité des clusters et possiblement changé les résultats.

5.1.6 Discussion sur les résultats

Les résultats des modèles avec et sans covariable ont permis de conclure à un effet de l'intervention dépendant du centre, c'est pourquoi les résultats ont été présentés de manière stratifiée. Cependant, même si les rapports de côtes du modèle final concernant le CHU de Nîmes et celui de Caen montrent une association significativement bénéfique de l'effet de l'intervention sur le taux d'erreurs médicamenteuses des IDE, le résultat pour le CHU de Bordeaux est radicalement inverse et, qui plus est, de manière significative.

Ceci peut être nuancé. En effet, comme dit dans le paragraphe ci-dessus, il aurait sans doute été préférable de modéliser une tendance temporelle dépendante du centre plutôt qu'un effet période global à tous les clusters. Comme constaté sur le descriptif de la variable réponse des données de l'étude, les variations des taux d'erreur avant et après l'intervention ne semblent pas suivre les mêmes tendances d'un centre à l'autre, ce qui n'est pas pris en compte par la modélisation. Précisément, le rapport de côte dans le modèle final comparant l'effet de la période deux par rapport à la période une est significativement positif. Ceci peut s'expliquer par l'augmentation importante du taux d'erreur à Bordeaux durant ces deux périodes d'observation avant l'intervention, sans que cette tendance soit partagée par le CHU de Caen (il est impossible de se prononcer sur le CHU de Nîmes qui reçoit l'intervention à la deuxième période d'observation). Cela pourrait expliquer une exagération de l'effet de l'intervention dans le CHU de Nîmes, par phénomène de compensation.

A l'inverse, le rapport de côtes dans le modèle final comparant l'effet de la période cinq par rapport à la période une est significativement négatif, ce qui dans ce cas s'explique sans doute par la différence du taux d'erreur entre la période quatre et la période cinq du CHU de Caen (correspondant aux périodes après l'intervention). Cela pourrait dans ce cas expliquer l'effet positif du rapport de côtes de l'effet de l'intervention au CHU de Bordeaux par même phénomène de compensation, puisqu'une tendance naturelle à la diminution du taux d'erreur est normalement attendue d'après les résultats du modèle. Ceci va dans le sens d'une confirmation dans le choix d'un effet temps aléatoire pour la modélisation des essais en stepped wedge plutôt qu'un effet traitement aléatoire. Effectivement, il peut être supposé dans le cas de l'étude OREANE que l'impossibilité de modélisation d'une tendance temporelle propre à chaque centre a introduit des biais potentiellement significatifs dans les résultats.

5.1.7 Les approches non paramétriques

Les approches non paramétriques peuvent être considérées comme des alternatives au modèle à intercept aléatoire dans le cas de comparaison verticale ou encore au modèle à pente aléatoire dans le cas de la comparaison verticale et horizontale. Cependant, ils ne permettent pas de prendre en compte la corrélation entre les données répétées au niveau individuel comme c'est le cas dans les stepped wedge de type cohorte.

L'utilisation de la méthode d'analyse intra-période sur les données de l'étude OREANE permet de comprendre pourquoi cette approche ne peut pas être utilisée dans ce cas. En effet, les données répétées sur les mêmes individus ont pour conséquence de diminuer la variance de la variable réponse. Ceci peut donc entraîner une mauvaise estimation de l'effet de l'intervention car celui-ci est estimé par l'intermédiaire d'une pondération inverse à la variance.

D'autre part, cette méthode utilise le test de permutation pour obtenir une p-value. Or, le nombre de permutations possibles, égale au nombre de cluster factoriel, doit être suffisamment important pour que cette dernière soit interprétable comme ce n'était pas le cas dans l'exemple de l'étude OREANE. A titre indicatif, si le nombre de clusters de l'étude OREANE avait été de six (comme étant recommandé au minimum), le nombre de permutations possibles serait de 720.

En outre, les analyses des approches non paramétriques entraînent une réduction des données disponibles. Par exemple, pour l'analyse intra-période, toutes les données où tous les clusters sont en condition soit de contrôle, soit d'intervention sont exclues.

Enfin, les approches non paramétriques ne permettent pas d'ajuster l'analyse sur différents facteurs comme c'est le cas avec les approches mixtes, ce qui explique pourquoi elles sont moins puissantes que ces dernières.

La méthode crossover qui est une approche faisant une comparaison verticale et horizontale n'a pas été utilisée comme exemple pour l'étude car elle réduit de manière très importante les données disponibles à l'analyse.

5.1.8 Les données manquantes de la variables réponse

Comme vu dans la partie du descriptif de la variable réponse, 6% sont des données manquantes. Le modèle de régression logistique à effets aléatoires est robuste dans le cas où les données sont manquantes aléatoirement (*missing at random* : MAR). A priori, cette hypothèse semble plausible. En effet, des variables observées qui expliquent le fait d'être manquant sont présentes dans le modèle (le centre et la période notamment). Par ailleurs, il ne paraît pas déraisonnable de penser qu'il n'existe pas de lien entre la valeur de la variable réponse et le fait que celle-ci soit manquante. En effet, la grande majorité des valeurs

manquantes concernent le CHU de Nîmes et celles-ci ont tendance à augmenter avec la période. Il est probable que le fait d'être manquant soit plutôt lié à un relâchement progressif dans le recueil de données au cours du temps, ce qui est un phénomène indépendant de la probabilité de faire une erreur.

5.2 Conclusion

5.2.1 Conclusion du stage et perspectives

La recherche bibliographique approfondie réalisée en début de stage a permis de dresser un bilan des principales méthodes statistiques d'analyse des stepped wedge en fonction des différents type de schéma (transversal et cohorte ouverte ou fermée). Les nombreux articles récents témoignent de l'évolution de la recherche dans ce domaine, même si certaines pistes restent encore à explorer notamment dans le cas des stepped wedge de type cohorte ouverte (rarement décrits dans la littérature à ce jour).

Par la suite, la modélisation statistique de l'étude OREANE a permis de répondre à la question principale en prenant en compte : les différents niveaux de corrélation, le potentiel effet de confusion du temps et diverses covariables d'ajustement. En revanche, le faible nombre de clusters disponibles dans l'étude ne permet pas de répondre à la question principale de manière généralisée à tous les centres et a impliqué des écarts aux méthodes classiques dans la stratégie de modélisation. Précisément, l'analyse statistique a montré grâce à un modèle de régression logistique mixte que l'effet de l'intervention était significativement dépendant du centre après ajustement sur les covariables sélectionnées. La formation complexe à l'étude a permis de diminuer le taux d'erreurs médicamenteuses dans le centre de Nîmes et de Caen, alors qu'un effet inverse a été observé dans le centre de Bordeaux. Par ailleurs, l'analyse de l'étude OREANE en tant que premier essai stepped wedge pris en charge par l'UMES-USMR, a permis de tirer des leçons pour les futures études suivant ce type de schéma : nombre de clusters minimum, méthodologie associée à ce schéma, stratégie de modélisation et calcul du nombre de sujets nécessaires.

Une des perspectives envisagée pour la suite du stage serait de modéliser l'erreur médicamenteuse à l'aide d'une régression de Poisson. En effet, comme beaucoup d'erreurs sont observées, cette variable peut être transformée pour obtenir des données de comptage. L'avantage de la régression de Poisson étant que les résultats s'interprètent en termes de risques relatifs, plus facilement interprétables par les cliniciens. Dans la continuité de mon stage, je vais m'intéresser à ce sujet en effectuant une revue approfondie de la littérature.

D'autre part, aucune méthode d'analyse non paramétrique pour les stepped wedge de type cohorte n'a été décrite à ce jour. Une recherche pourra être effectuée dans le but de savoir si des méthodes non paramétriques développées dans d'autres contextes permettent

de prendre en compte les données corrélées au niveau individuel et si ces dernières pourraient s'appliquer dans le cadre des stepped wedge de type cohorte. Le cas échéant, de telles méthodes pourraient être testées sur les données de l'étude OREANE à titre illustratif.

5.2.2 Conclusion personnelle du stage

Le stage de Master 2 s'est déroulé au sein de l'USMR où j'ai exercé le métier de data manager pendant plusieurs années. Ceci a eu pour conséquences de faciliter mon adaptation en tant que stagiaire car que je connaissais déjà la plupart des membres de l'équipe, même si le stage s'est déroulé intégralement à distance jusqu'à présent.

Au cours de ce stage, j'ai beaucoup appris sur les différents aspects du métier de biostatisticien en analysant les données de l'étude OREANE. Par ailleurs, j'ai également été intégré à un groupe de travail « stepped wedge » initié par l'UMES. Ceci m'a permis de participer à des réunions et consultations méthodologiques de différents projets en construction. Hormis les aspects d'analyse statistique, j'ai donc également pu acquérir de l'expérience dans le travail de réflexion et de préparation d'un projet de recherche, ce qui est une étape très intéressante et indispensable au métier de biostatisticien. Aussi, j'ai également participé aux réunions de l'UMES qui travaille en collaboration avec l'USMR. Je me suis senti pleinement intégré à l'USMR mais également dans son environnement de travail collaboratif.

Egalement, ce stage a nécessité un travail de recherche important sur les méthodes d'analyse statistique des essais en stepped wedge. Le travail bibliographique réalisé m'a apporté une connaissance à la fois sur ce schéma d'étude, mais également sur les essais randomisés en clusters. Cela m'a également permis de consolider mes connaissances dans les méthodes d'analyse de données longitudinales et de découvrir d'autres méthodes non paramétriques.

6 REFERENCES BIBLIOGRAPHIQUES

Barker, D., McElduff, P., D'Este, C., and Campbell, M.J. (2016). Stepped wedge cluster randomised trials: A review of the statistical methodology used and available. BMC Med. Res. Methodol. *16*.

Barker, D., D'Este, C., Campbell, M.J., and McElduff, P. (2017). Minimum number of clusters and comparison of analysis methods for cross sectional stepped wedge cluster randomised trials with binary outcomes: A simulation study. Trials *18*, 1–11.

Hemming, K., Haines, T.P., Chilton, P.J., Girling, A.J., and Lilford, R.J. (2015a). The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. BMJ *350*, h391.

Hemming, K., Lilford, R., and Girling, A.J. (2015b). Stepped-wedge cluster randomised controlled trials: A generic framework including parallel and multiple-level designs. Stat. Med. *34*, 181–196.

Hemming, K., Taljaard, M., and Forbes, A. (2017). Analysis of cluster randomised stepped wedge trials with repeated cross-sectional samples. Trials *18*, 101.

Hemming, K., Kasza, J., Hooper, R., Forbes, A., and Taljaard, M. (2020). A tutorial on sample size calculation for multiple-period cluster randomized parallel, cross-over and stepped-wedge trials using the shiny CRT calculator. Int. J. Epidemiol. *49*, 979–995.

Hooper, R., and Bourke, L. (2015). Cluster randomised trials with repeated cross sections: Alternatives to parallel group designs. BMJ *350*.

Hussey, M.A., and Hughes, J.P. (2007). Design and analysis of stepped wedge cluster randomized trials. Contemp. Clin. Trials *28*, 182–191.

Kennedy-Shaffer, L., de Gruttola, V., and Lipsitch, M. (2020). Novel methods for the analysis of stepped wedge cluster randomized trials. Stat. Med. *39*, 815–844.

Li, F., Turner, E.L., and Preisser, J.S. (2018). Sample size determination for GEE analyses of stepped wedge cluster randomized trials. Biometrics *74*, 1450–1458.

Mdege, N.D., Man, M.S., Taylor, C.A., and Torgerson, D.J. (2011). Systematic review of stepped wedge cluster randomized trials shows that design is particularly used to evaluate interventions during routine implementation. J. Clin. Epidemiol. *64*, 936–948.

Thompson, J.A., Fielding, K., Hargreaves, J., and Copas, A. (2017). The optimal design of stepped wedge trials with equal allocation to sequences and a comparison to other trial designs. Clin. Trials *14*, 639–647.

Thompson, J.A., Davey, C., Fielding, K., Hargreaves, J.R., and Hayes, R.J. (2018). Robust analysis of stepped wedge trials using cluster-level summaries within periods. Stat. Med. *37*, 2487–2500.

Voldal, E.C., Hakhu, N.R., Xia, F., Heagerty, P.J., and Hughes, J.P. (2020). swCRTdesign: An R Package for Stepped Wedge Trial Design and Analysis. Comput. Methods Programs Biomed. *196*, 105514.