

Table des matières

Remerciements	i
Table des matières	iii
Table des figures	iv
Liste des tableaux	v
Liste des abréviations	vi
Introduction	1
1 CONTEXTE ET GENERALITES	3
1.1 Contexte de l'étude	3
1.1.1 L'unité d'Épidémiologie et de Recherche Clinique (EPI-RC) de l'IPM	3
1.1.2 L'USAID et les programmes Mikolo et Mahefa Miaraka	3
1.1.3 Le projet «Étude des Déterminants du Recours aux Soins des femmes enceintes et des enfants de moins de 5 ans»(EDRaS)	4
1.1.4 Présentation et description du travail de mémoire	4
1.2 Généralités	5
1.2.1 La consultation prénatale	5
1.2.2 Rappels théoriques sur les tests statistiques et sur les modèles de régression	5
2 MATERIELS ET METHODES	12
2.1 Site de l'étude	12
2.2 Population d'étude	14
2.3 Analyse de données	16
2.3.1 Outils	16
2.3.2 Préparation des données	16
2.3.3 Analyses biostatistiques	17
3 RESULTATS, INTERPRETATIONS et DISCUSSIONS	21
3.1 Inclusion de la population	21
3.2 Recours aux CPN	23
3.3 Description de l'échantillon	25

3.4	Analyse des facteurs de recours aux CPN	29
3.4.1	Analyses uni variées	29
3.4.2	Analyses multivariées : Modèles obtenus	30
3.5	Évaluation des méthodes	33
3.5.1	<i>LOGITS</i> Adjacents et Cumulés	33
3.5.2	Modèle à <i>LOGITS</i> Cumulés : Comparaison des deux résultats avec effets aléatoires districts et communes	35
Conclusion		36
Bibliographie		39
Annexes		I

Table des figures

1.1	Courbe logistique	7
2.1	Site d'étude et sélection des Fokontany	13
3.1	Flux d'inclusion des femmes dans le cadre du projet EDRaS, IPM 2017-2018	22
3.2	Proportion des femmes par classe du recours aux CPN	23
3.3	Classe de Recours aux CPN par commune	24
3.4	Distribution des femmes dans chaque district selon les classes d'âge. Résultat EDRaS 2018	25
3.5	Grossesse non désirée. Source Projet EDRaS- IPM 2017-2018	II
3.6	Diagramme de flux d'inclusion des Fokontany. Source Projet EDRaS- IPM 2017-2018	III
3.7	Carte des Fokontany inclus dans le district de Moramanga. Source Projet EDRaS- IPM 2017-2018	IV
3.8	Carte des Fokontany inclus dans le district de Manakara. Source Projet EDRaS- IPM 2017-2018	V
3.9	Carte des Fokontany inclus dans le district d'Antsohihy. Source Projet EDRaS- IPM 2017-2018	VI

Liste des tableaux

2.1 Critère d'inclusion	15
2.2 Gestion des données manquantes	16
3.1 Association entre le nombre de CPN et les districts	24
3.2 Distributions des profils sociodémographiques selon les classes de recours aux CPN. Analyse descriptive - Résultat EDRaS 2018	26
3.3 Distributions des profils de santé selon les classes de recours aux CPN. Analyse descriptive - Résultat EDRaS 2018	27
3.4 Distributions des perceptions femmes et le rôle de leurs entourages selon les classes de recours aux CPN. Analyse descriptive - Résultat EDRaS 2018	28
3.5 Association entre les profils sociodémographiques, profils de santé et la perception des femmes avec leurs recours aux soins prénatals. Analyses uni variées.	29
3.6 Facteurs favorisant le passage de niveau en niveau en nombre de séances de CPN effectuées par les femmes. <i>LOGITS</i> adjacents. Résultat EDRaS 2018	31
3.7 Facteurs d'augmentation du nombre de séances de CPN chez les femmes enceintes sous l'effet aléatoire : District. <i>LOGIT</i> cumulé. Résultat EDRaS 2018	32
3.8 Facteur d'augmentation du nombre de séances de CPN chez les femmes enceintes sous l'effet aléatoire : Commune. <i>LOGIT</i> cumulé. Résultat EDRaS 2018	32
3.9 Tableau comparatif des résultats des deux méthodes du modèle logistique ordinaire	33
3.10 Matrice de confusion obtenu par le modèle à <i>LOGITS</i> adjacents	34
3.11 Matrice de confusion obtenu par le modèle à <i>LOGITS</i> cumulés	34

Liste des abréviations

ANOVA	Analysis Of Variance
AIC	Akaike Information Criterion
CPN	Consultation Périnatale
EDRaS	Etudes des Déterminants du Recours aux Soins des mères et enfants de moins de 5 ans
ENSOMD	Enquête Nationale sur le Suivi des Objectifs du Millénaire pour le Développement à Madagascar
IPM	Institut Pasteur de Madagascar
OMS	Organisation Mondiale de la Santé
ONG	Organisation Non Gouvernementale
OR	<i>Odds – Ratio</i>
SA	Semaine d'Aménorrhée
SGBD	Système de Gestion de Base de Données
USAID	United States Agency for International Development

Introduction

Au niveau mondial, la santé maternelle et infantile reste un enjeu de santé publique majeur. Selon l’OMS^[1], 830 femmes décèdent chaque jour pour des causes liées à la grossesse ou à l’accouchement. La majorité de ces décès sont enregistrés dans les pays en voie de développement et apparaissent surtout en milieu rural. A Madagascar, on enregistre 478 décès maternels et 26 décès infantiles pour 100 000 naissances vivantes selon ENSOMD^[10]. Ces situations sont notamment liées à une prise en charge insuffisante, tardive ou à une faible utilisation des services de soins existants. Or, il a été démontré que le recours aux soins préventifs et curatifs de qualité peut sauver la vie des femmes et de leurs enfants et éviter la survenue de maladies ou de complications.

Les consultations prénatales (CPN) sont reconnues comme étant l’un des trois principaux piliers de la lutte contre la mortalité maternelle et infantile^[2]. L’intérêt porté aux soins prénatals résulte du fait que la santé de la mère engage aussi celle de son futur enfant^[19]. Pour être efficaces, les soins prénatals doivent se poursuivre avec une certaine régularité jusqu’à l’accouchement. L’Organisation Mondiale de la Santé (OMS) recommande au moins 4 visites prénatales tout au long de la grossesse. Or, dans notre pays, seulement 51% des femmes enceintes ont effectué 4 visites prénatales ou plus en 2013^[3].

A notre connaissance, il y a peu d’études récentes sur la prévalence et les facteurs du recours aux CPN chez les femmes enceintes à Madagascar. Pourtant, des études similaires pourraient améliorer les stratégies de réduction de la mortalité materno-infantile dans le pays.

L’étude sur la détermination des facteurs de recours aux soins vise donc à trouver un modèle statistique (ou équation mathématique) qui permet d’expliquer les variations du recours aux soins (dite variable dépendante) en fonction des variables explicatives sélectionnées. Il est toujours nécessaire d’améliorer le modèle statistique ainsi obtenu afin d’obtenir un résultat fiable et adapté à l’étude. Pour une variable dépendante ordinale, différentes méthodes peuvent être utilisées pour améliorer d’une part l’interprétabilité des résultats du modèle. D’autre part, un aspect, souvent non considéré, pourrait aussi affiner un modèle statistique. Il s’agit d’étudier les variabilités liées à la géographie ou à l’appartenance d’un individu à un groupe donné ; c’est-à-dire, de prendre ce groupe comme effet aléatoire au modèle. Pour bien mettre en cause ces variabilités, il convient de bien choisir le groupe (ou l’unité géographique) à utiliser dans le modèle.

Cette étude vise à cerner les facteurs de recours en consultation prénatale pendant la période de grossesse des femmes, vivant en milieux ruraux dans les districts d’ANTSOHIHY,

-
1. <https://www.who.int/fr/news-room/fact-sheets/detail/maternal-mortality>
 2. Les 2 autres piliers : PF et accouchement dans les conditions optimales d’hygiène et santé
 3. Millenium Development Goals Indicators - 2013

de MORAMANGA et de MANAKARA. Principalement, elle permet d'estimer l'utilisation des soins prénatals dans ces districts. Pour ce faire, nous modélisons les chances (ou probabilité) pour une femme de compléter ou pas ses séances de CPN durant sa grossesse. Deux méthodes de la régression logistique polytomique ordinale sont étudiées et évaluées. Secondairement, les variabilités liées aux unités géographiques (districts et communes) sont prises en compte dans les analyses en introduisant l'effet mixte au modèle.

Pour atteindre nos objectifs, ce travail comprend essentiellement trois parties : la première partie est consacrée à la présentation du contexte global ainsi que la description de l'étude. La deuxième partie décrit les matériels et méthodes utilisés. La dernière partie présente les résultats ainsi obtenus, leurs interprétations respectives et les discussions.

Chapitre 1

CONTEXTE ET GENERALITES

1.1 Contexte de l'étude

Ce travail de mémoire a été effectué au sein de l'unité d'Épidémiologie et de Recherche Clinique de l'Institut Pasteur de Madagascar, dans le cadre d'un projet multidisciplinaire EDRaS (Étude des Déterminants du Recours aux Soins). Ce dernier a été financé par l'USAID (United States Agency for International Development), par l'intervention des programmes Mikolo et Mahefa Miaraka.

1.1.1 L'unité d'Épidémiologie et de Recherche Clinique (EPI-RC) de l'IPM

L'unité d'Épidémiologie et de Recherche Clinique fait partie des neuf unités de recherche de l'IPM. Elle comporte un groupe Santé Géomatique (SAGEO), un groupe Santé et Science Sociale (SASS) et un Groupe de Biostatistique et Bio-mathématiques (G2B) dans lequel ce stage a été rattaché. Ses activités principales sont les études cliniques et épidémiologiques, les études socio-anthropologiques des comportements et pratiques en santé, l'appui au Ministère de la Santé Publique dans les investigations des épidémies, le renforcement des compétences des médecins de la santé publique en matière de la surveillance des maladies à potentiel épidémique et aussi l'accueil et l'encadrement de ses stagiaires.

1.1.2 L'USAID et les programmes Mikolo et Mahefa Miaraka

L'USAID ou United States Agency for International Development est une agence indépendante du gouvernement des Etats-Unis chargée du développement économique et de l'assistance humanitaire dans le monde. Elle travaille à Madagascar depuis une trentaine d'années pour aider la population malagasy à réaliser ses objectifs de développement en dépit d'une conjoncture sociale difficile. Ses activités sont principalement l'amélioration de la qualité des soins, l'aide aux agriculteurs, des projets environnementaux, et l'aide en cas de catastrophe, etc^[1].

1. USAID : <https://www.usaid.gov/fr/madagascar/our-work>

Pour coordonner ces activités à Madagascar, l'USAID met en place des programmes essentiellement communautaires dont Mikolo et Mahefa Miaraka font partie. Le projet Mikolo a été créé depuis 2013 et s'est terminé en 2018. Il visait à réduire la morbidité et la mortalité maternelle et infantile en augmentant l'utilisation des services de soins de santé primaires communautaires et en encourageant les femmes et les enfants à adopter des comportements sains. Ce projet intervenait dans 9 régions de Madagascar incluant les districts de MORAMANGA et de MANAKARA.

Outre, le programme Mahefa Miaraka vise à renforcer la capacité du secteur public, à planifier, à approvisionner, à gérer les services de santé communautaire et à relancer l'engagement communautaire dans le domaine de la santé. De ce fait, il contribue à améliorer les services de santé de base dans 7 régions et 34 districts, y compris le district d'ANTSOHIHY.

1.1.3 Le projet «Étude des Déterminants du Recours aux Soins des femmes enceintes et des enfants de moins de 5 ans» (EDRaS)

Le projet EDRaS, sous le financement de l'USAID, a été mené par l'Institut Pasteur de Madagascar par le biais de l'unité d'Epidémiologie et de Recherche Clinique de l'institut. Il vise à déterminer les :

- Recours aux soins (préventifs et curatifs) durant la grossesse ;
- Recours aux établissements de soins ou à des personnels de soins qualifiés pour l'accouchement ;
- Recours à l'utilisation d'une méthode de planification familiale (avant et après la grossesse) ;
- Recours aux soins en période néo-natale (0 à 28 jours) ;
- Recours aux soins (préventifs et curatifs) pour les enfants de moins de 5 ans.

1.1.4 Présentation et description du travail de mémoire

La présente étude s'est portée sur la partie quantitative du projet EDRaS. Elle consiste donc à évaluer les facteurs influençant le recours ou non aux consultations prénatales et à estimer l'utilisation des soins prénatals dans les trois districts (Antsohihy, Moramanga et Manakara).

Sous ce contexte, les méthodologies et la collecte de données se sont faites sous les protocoles du projet EDRaS, mais le travail de mémoire permet d'apporter plus de profondeur sur l'exploitation des données collectées.

En effet, deux points sont mis en valeur dans ce travail :

- Nous ne classifions pas seulement le recours aux soins prénatals par avoir recours ou non, mais nous le qualifions selon le nombre de séances effectuées. Les résultats ainsi obtenus seront plus informatifs et répond mieux à l'objectif du projet EDRaS.

- Nous prenons en compte les effets des unités géographiques en faisant l’hypothèse que la géographie a une influence différente sur ce type de recours.

Pour trouver les méthodes adaptées aux objectifs, nous utilisons deux approches de la régression ordinale : à *LOGITS* adjacents et à *LOGITS* cumulatifs à effet mixte.

1.2 Généralités

Dans cette partie, nous allons voir les définitions et descriptions de la CPN, des tests statistiques et des modèles de régression.

1.2.1 La consultation prénatale

La consultation prénatale est l’ensemble des éléments de suivi d’une femme enceinte pendant sa grossesse. Elle permet de dépister, de prévenir ou de prendre en charge certaines affections susceptibles de menacer la santé de la mère et de l’enfant. Elle permet aussi de conseiller, de soutenir et d’informer la femme enceinte ainsi que sa famille avant l’accouchement. Les agents qui peuvent effectuer la CPN sont des personnels de santé qualifiés (médecin, sage-femme, infirmier formé). En bref, le but des soins prénatals est d’assurer la santé et l’accouchement à moindre risque.

Elle est classée comme l’une des trois principaux piliers de la lutte contre la mortalité materno-infantile ; les deux autres étant la planification familiale et les soins obstétricaux dont l’accouchement dans les conditions optimales d’hygiène et de sécurité.

Une étude sur le comportement maternel vis-à-vis de la surveillance prénatale a démontré qu’il faut mettre l’accent sur l’amélioration des offres de soins prénatals afin de prévenir les mortalités périnatales [19].

L’Organisation Mondiale de la Santé (OMS) recommande au moins 4 visites prénatales à intervalles réguliers tout au long de la grossesse (au premier, deuxième, troisième trimestre de grossesse et juste avant l’accouchement) [2].

1.2.2 Rappels théoriques sur les tests statistiques et sur les modèles de régression

Test statistique

Un test statistique est une procédure de décision entre deux hypothèses concernant un ou plusieurs échantillons [14]. L’hypothèse nulle notée H_0 est celle que l’on considère vraie à priori tandis que l’hypothèse alternative H_1 est l’hypothèse complémentaire de H_0 . Le but du test

2. Actuellement c’est 8 séances mais nous nous sommes restés sur l’ancienne recommandation pour cette étude.

est de décider si cet à priori est crédible : H_1 est choisie uniquement par défaut si H_0 n'est pas considérée comme crédible (WikiStat). Le choix de ces deux hypothèses est en général imposé par le test qu'on utilise. Parmi ces tests il y a le test de Chi2, le test exact de Fisher, le test de Student et l'ANOVA.³

La régression logistique

Faisant partie de l'apprentissage supervisé, la régression logistique est une technique de modélisation qui vise à prédire et à expliquer les valeurs d'une variable catégorielle (dite variable dépendante ou variable à expliquer) à partir d'une variable ou d'une collection de variables continues ou catégorielles⁴.

Elle est dite binaire lorsque la variable dépendante est à deux modalités (en terme de santé, c'est souvent l'apparition ou non d'une maladie)⁵. Elle est dite multinomiale lorsque la variable dépendante est à plus de deux modalités. La régression logistique multinomiale est souvent utile dans le cas où nous souhaitons classer des objets en fonction des valeurs d'un groupe de variables de prédicteur. Ce type de régression est similaire à la régression logistique binaire, mais s'avère plus général puisque la variable dépendante n'est pas seulement limitée à deux catégories¹⁷.

Elle est dite ordinale si les modalités de la variable à expliquer se présentent dans un ordre hiérarchique¹¹. Comme exemple, nous pouvons classer l'état d'une maladie en : moins grave – grave – sévère.

Approche statistique et probabiliste de la régression logistique binaire

Soit Y la variable aléatoire à prédire (à 2 modalités $\{1, 0\}$). Soient $X = (X_1, X_2, \dots, X_p)$ p variables explicatives. Le principe de la régression logistique est de modéliser la probabilité conditionnelle $P(Y = 1|X_1, X_2, \dots, X_p)$, notée $\pi(X)$.

Cette probabilité n'est que la moyenne :

$$E(Y/X) = 1 \times P(Y = 1/X) + 0 \times P(Y = 0/X) = P(Y = 1|X) = \pi(X).$$

, communément appelée «fonction logistique».

Dans le cas d'une seule variable explicative X , la fonction logistique s'écrit par définition :

$$\pi(X) = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}}$$

La courbe logistique a ainsi une forme sigmoïde :

3. Analysis Of Variance

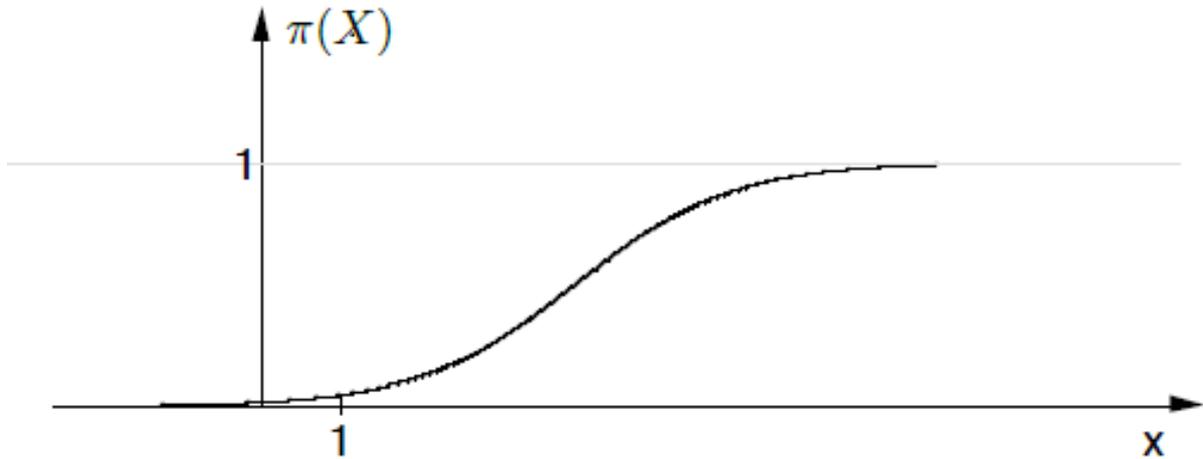


FIGURE 1.1 – Courbe logistique

Dans le cas de plusieurs variables explicatives $X = (X_1, X_2, \dots, X_p)$ ($p > 1$), la fonction logistique $\pi(X)$ se généralise sous la forme :

$$\pi(X) = \frac{e^{\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p}}{1 + e^{\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p}} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p)}} \quad (1.1)$$

– On vérifie que $\pi(X) \in [0, 1]$ car si X tend vers $-\infty$, $\pi(X)$ tend vers 0. Et si X tend vers $+\infty$, $\pi(X)$ tend vers 1.

– $\beta_0, \beta_1, \dots, \beta_p$ sont les paramètres à estimer.

En utilisant la transformation *LOGIT* : $Logit(\nu) = \ln\left(\frac{\nu}{1-\nu}\right)$, avec l'équation (1.1), nous avons

$$Logit(\pi(X)) = \ln\left(\frac{\pi(X)}{1-\pi(X)}\right) = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p \quad (1.2)$$

Ce qui fait apparaître le modèle logistique comme un modèle linéaire [5].

Un intérêt majeur de la régression logistique réside dans la notion de l'**odds-ratio** (interprétation des coefficients du modèle comme une mesure d'association des X_i par rapport aux modalités de Y).

1. Estimation des paramètres

Sachons qu'en régression linéaire multiple, on estime les coefficients par la méthode des moindres carrés ordinaires (MCO). Mais cette méthode ne permet pas d'estimer les coefficients

en régression logistique. D'où l'utilisation de la méthode du maximum de vraisemblance. Supposons que nous n'avons qu'une seule variable explicative X .

La vraisemblance de l'observation (X, Y_i) est, par définition, la probabilité d'observer cette observation, considérée comme une fonction des paramètres inconnus β_0 et β_1 .

La vraisemblance (notée L) pour $Y_i = 1$ est $L_1 = \frac{1}{e^{-(\beta_0 + \beta_1 X)}}$ et $L_0 = 1 - L_1 = \frac{e^{-(\beta_0 + \beta_1 X)}}{1 + e^{-(\beta_0 + \beta_1 X)}}$

La vraisemblance L de l'ensemble des observations (probabilité d'observer l'ensemble de l'échantillon) est le produit des probabilités L_1 et L_0 (en supposant que les observations sont statistiquement indépendantes). On obtient

$$L = \frac{\prod_n e^{-(\beta_0 + \beta_1 X)}}{\prod_N 1 + e^{-(\beta_0 + \beta_1 X)}}$$

où n le nombre d'observations dont $Y = 1$ et N le nombre total des observations [5].

La méthode du maximum de vraisemblance permet donc de maximiser cette probabilité en choisissant les $\hat{\beta}_0$ et $\hat{\beta}_1$ qui la rendent maximum. Pour ce faire, on utilise le Log-vraisemblance (permet de faciliter le calcul) car les β maximisant $\text{Log}(L)$ maximisent aussi L d'après la propriété de la fonction logarithme (monotone croissante). Il n'y a pas de solutions explicites pour β que dans le cas d'un seul coefficient (une seule variable explicative). Sinon on procède par itération et les valeurs obtenues sont des approximations numériques [11].

2. Interprétation des coefficients

odds-ratio ou rapport des côtes :

C'est le rapport des côtes de deux valeurs X_i et X_j d'une variable explicative X

$$OR = \frac{\frac{P(Y=1|X=X_i)}{P(Y=0|X=X_i)}}{\frac{P(Y=1|X=X_j)}{P(Y=0|X=X_j)}}$$

L' OR s'interprète comme : la probabilité pour que $Y = 1$ au lieu de 0 est OR fois plus forte pour $X = X_i$ que pour $X = X_j$.

L'*odds - ratio* liée à une modalité X_i de X est $e^{\beta x_i}$ [13]. En effet, considérons X comme une seule variable explicative binaire ($X = 1$ exposé et $X = 0$ non exposé), d'après (1.2), nous avons :

$$\text{LOGIT}(\pi(X)) = \beta_0 + \beta_1 X$$

Ensuite, $\text{LOGIT}(\pi(1)) = \beta_0 + \beta_1$ et $\text{LOGIT}(\pi(0)) = \beta_0$.

Ainsi, nous avons : $\text{LOGIT}(\pi(1)) - \text{LOGIT}(\pi(0)) = \ln\left(\frac{\frac{\pi(1)}{1-\pi(1)}}{\frac{\pi(0)}{1-\pi(0)}}\right) = \ln(OR) = \beta_1$

Et nous avons par conséquent, $OR = e^{\beta_1}$

De même pour le cas à plusieurs variables explicatives, e^{β_k} est l'*odds - ratio* liée à une modalité X_k d'une variable explicative X ajusté sur les autres variables.

La régression logistique multinomiale (ou régression logistique polytomique nominale)

Comme dans la régression logistique binaire, on dispose d'une variable dépendante Y mais à m modalités (m étant supérieur à 2), et des variables explicatives $X = (X_1, \dots, X_p)$. Les modalités de Y sont supposées non ordonnées ou sans lien hiérarchique.

Dans la pratique, ce modèle permet de modéliser l'appartenance d'un individu à une catégorie (modalité) k : $\pi_k(X) = P(Y = u_k|X)$.

L'approche consiste alors à se donner une modalité de référence u_{ref} de Y , et pour tout $k \in \{1, \dots, u_{ref}, \dots, u_m\}$, nous avons la réécriture du *LOGIT* :

$$LOGIT(\pi_k) = \ln\left(\frac{\pi_k}{\pi_{ref}}\right) = \ln\left(\frac{P(Y = u_k|X)}{P(Y = u_{ref}|X)}\right) = \beta_{0,k} + \beta_{1,k}X_1 + \dots + \beta_{p,k}X_p \quad (1.3)$$

où $\beta_{0,k}, \beta_{1,k}, \dots, \beta_{p,k}$ des coefficients réels inconnus (relatifs à la modalité u_k) que l'on cherche à estimer.

Pour chaque modalité u_k ($k \neq ref$),

$$\pi_k = P(Y = u_k|X) = \frac{\exp(\beta_{0,k} + \beta_{1,k}X_1 + \dots + \beta_{p,k}X_p)}{1 + \sum_{k=2}^m \exp(\beta_{0,k} + \beta_{1,k}X_1 + \dots + \beta_{p,k}X_p)} \quad (1.4)$$

Et on en déduit $\pi_{ref} = 1 - \sum_{k \neq ref} \pi_k$

Lorsque $k = 2$, $u_{ref} = u_1 = 0$ et $u_k = u_2 = 1$, alors cela correspond au modèle de régression logistique binaire.

Quand on estime donc les coefficients $\beta_{0,k}, \beta_{1,k}, \dots, \beta_{p,k}$, on pourra estimer $P(Y = u_k|X)$ par substitution. L'estimation se fait en utilisant le maximum de vraisemblance (on s'appuie sur la loi multinomiale pour écrire cette vraisemblance [\[18\]](#)).

La régression logistique polytomique ordinaire

Moins connue, la régression logistique polytomique ordinaire est une régression logistique où les modalités de la variable dépendante Y sont ordonnées [\[3\]](#). Le fait qu'une modalité $k + 1$ de Y succède k a du sens.

Pour ce type de régression, nous considérons deux approches telles que l'idée de *LOGITS* adjacents et de *LOGITS* cumulés [15].

Soit Y la variable dépendante désignant l'état d'une maladie, codé en trois niveaux de modalités : non grave, grave et sévère.

1. Le modèle à *LOGITS* adjacents

Ce modèle permet de modéliser le *LOGIT* du passage d'un niveau à l'autre qui lui est adjacent avec une combinaison linéaire des variables explicatives. Dans notre exemple, c'est la modélisation de l'aggravation d'une maladie (de «non grave» à «grave» ou de «grave» à «sévère»).

2. Le modèle à *LOGITS* cumulés

Cette méthode permet de comparer un niveau avec tous les niveaux qui lui sont inférieurs (ou supérieur). Donc pour chaque niveau k , on modélise le *LOGIT* quantifiant le surcroît de chances d'être en dessous plutôt qu'en dessus de ce niveau (ou l'inverse).

Le modèle à effet mixte

On utilise un modèle à effet fixe (ou modèle de régression) lorsqu'on veut expliquer une variable dépendante Y à partir des variables explicatives sans tenir compte de l'aspect d'appartenance d'un individu à un groupe donné ou d'autre aspect qui pourrait causer le fait que des observations différentes partagent un même «contexte».

Avec un modèle à effet fixe, on veut étudier les variations de la variable dépendante Y en fonction des variables explicatives X . Mais, avec les effets mixtes, on suppose de plus que d'autres caractéristiques communes inobservées sont portées par plusieurs observations, également corrélées avec les variations de Y [2].

Prenons le modèle linéaire simple pour expliquer le modèle à effets mixtes. Soit Y la variable d'intérêt désignant les notes des élèves d'une école. En modèle linéaire simple, on suppose une dépendance linéaire entre les notes et l'âge des élèves, avec un terme d'erreur ε , soit l'équation :

$$Y = \alpha + \beta.x + \varepsilon \tag{1.5}$$

où x désigne les notes, $\varepsilon \sim N(0, \sigma^2)$ terme d'erreur, α , β et σ paramètres à estimer.

Nous avons considéré que tous les élèves ont le même comportement quel que soit sa classe, et que les erreurs sont indépendantes. Or, il se pourrait que les notes des élèves d'une même classe présentent un certain degré de dépendance (ou corrélation) comme la gentillesse de l'instituteur.

Pour prendre en compte cet aspect "classe" (dépendance entre élèves d'une même classe), on rajoute à notre modèle à effet fixe un effet aléatoire. On obtient ainsi un modèle à effets

mixte c'est-à-dire qui possède à la fois un ou des effets fixes et un ou des effets aléatoires. Les équations ci-dessous nous montrent le modèle(mixte) où l'on ajoute une constante aléatoire (1.6) et le modèle mixte à pente aléatoire(1.7) [6] :

$$Y_j = (\alpha + \alpha_j) + \beta.x + \varepsilon_j \quad (1.6)$$

Dans ce cas, les élèves ont une note initiale moyenne de α quelque soit leur âge, mais cette note varie de α_j pour les élèves d'une même classe j . Notons que $\alpha_j \sim N(0, \mu^2)$.

$$Y_j = \alpha + (\beta + \beta_j).x + \varepsilon_j \quad (1.7)$$

La note des élèves varient en moyenne de β en fonction de l'âge, mais il y a une déviation de β_j par rapport à cette moyenne pour les élèves d'une même classe. Et $\beta_j \sim N(0, \nu^2)$

Un modèle à effets aléatoires est conçu pour rendre compte de composantes «inobservée» du modèle⁴

4. <https://statoscope.wordpress.com/2017/06/25/donnees-groupees-effets-aleatoires-vs-effets-fixes/>

Chapitre 2

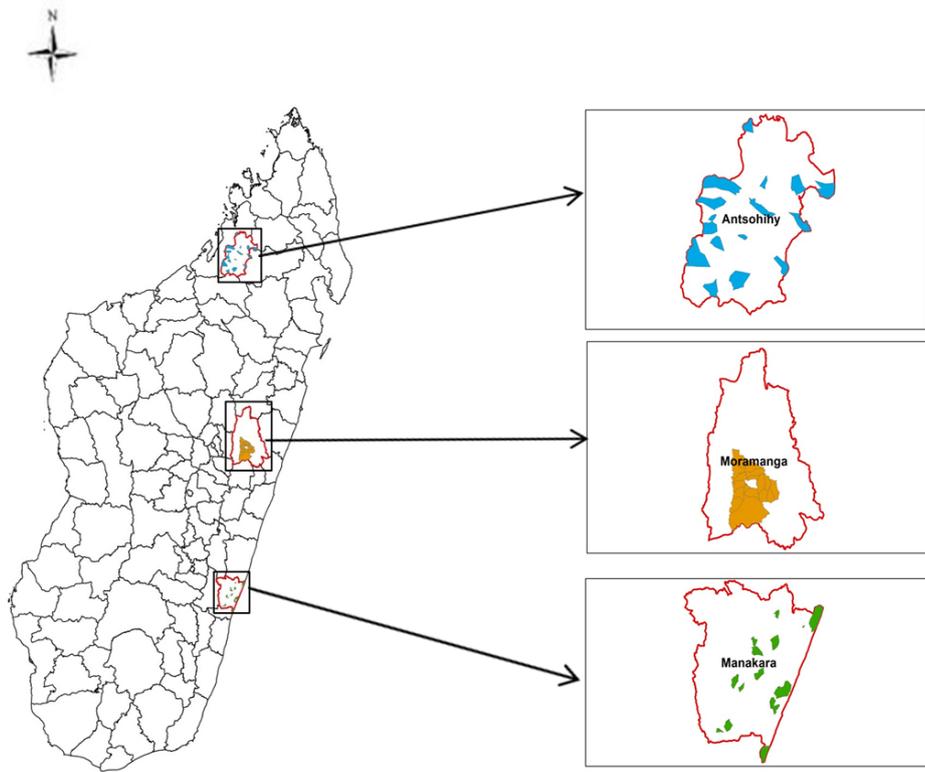
MATERIELS ET METHODES

2.1 Site de l'étude

Dans son intégralité, l'étude s'est déroulée dans 03 districts appuyés par les programmes Mikolo et/ou Mahefa Miaraka. Ils appartiennent tous à la catégorie des régions les mieux desservies en CSB public : le ratio CSB public sur la population varie de 1/7 621 à 1/8 999 personnes^[9]

- Le district d'Antsohihy dans la région de Sofia est une des zones d'intervention de l'ONG Mahefa-Miaraka. Ce district est à 700 km, au Nord-Ouest de la capitale sur la route vers la région Diana. C'est une zone à faible altitude (19 m), et la population était au nombre de 144 970 habitants en 2015. L'ONG Mahefa Miaraka couvre tout le district.
- Le district de Moramanga dans la région Alaotra Mangoro, est situé à 115km à l'Est de la capitale vers la route de Toamasina. L'ONG Mikolo y est présente en partie. Dans ce district, l'IPM a mis en place un observatoire démographique et de santé des populations couvrant 3 communes dont : la Commune Urbaine de Moramanga, Ambohibary et Ampasimpotsy. En 2015, le district incluait 290 236 habitants. La présence de l'Observatoire facilitera l'accès aux diverses informations.
- Dans la partie Sud Est, le district de Manakara appartenant à la région de Vatovavy Fitovinany situé en zones côtières abordant l'Océan Indien à l'Est, est une zone appuyée par l'ONG Mikolo, qui est plus présent en milieu rural. Le canal des Pangalanes traverse ce district et la population du district se chiffrait à 385 468^[1] en 2015.

1. Données de projection de population de district INSTAT-Serveur SIG –Unité Epidémiologie IPM



Légende:

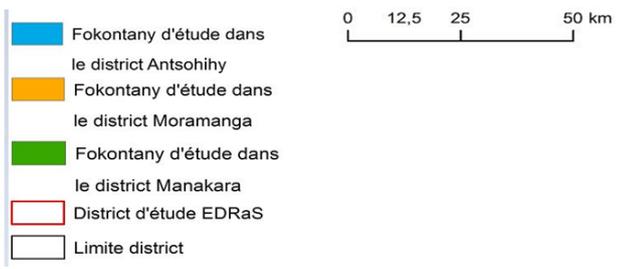


FIGURE 2.1 – Site d'étude et sélection des Fokontany

2.2 Population d'étude

Nous avons fait une étude observationnelle en communauté chez les femmes enceintes et mères d'enfants de moins de 5 ans avec 2 passages : un pendant la grossesse et un à partir de 1 mois et demi après l'accouchement.

Pour cette étude, le Fokontany a été l'unité de randomisation vu que dans leurs communes d'interventions respectives, les ONG Mikolo et Mahefa-Miaraka interviennent à l'échelle des Fokontany.

Les résultats préliminaires au niveau de l'observatoire de suivi de population à Moramanga ont montré 2 à 3% de proportion de femmes enceintes. Par souci de faisabilité, la proportion de 2% est retenue afin de maximiser les chances de trouver et d'inclure le nombre de sujet nécessaires en un temps limité.

En tenant compte de ces proportions, les Fokontany devraient avoir au moins 1050 habitants pour être éligible.

Un tirage au sort des Fokontany répondant au critère d'éligibilité a été alors effectué et les huit premiers Fokontany, par zone d'étude, tirés au sort ont été inclus dans l'étude.

Toutes les femmes remplissant les critères d'inclusions et habitant dans ces Fokontany ont été invitées à participer à l'étude de manière exhaustive.

TABLE 2.1 – Critère d’inclusion

Population	Critère
Population cible	Femmes enceintes (et/ou mères d’enfants de moins de 5 ans)
Population source	Toutes les femmes enceintes des Fokontany choisis dans les 3 sites d’études
Critères d’inclusion des Femmes enceintes	Femmes enceintes de 24 SA ou plus (à partir du 6ème mois de grossesse), Résident permanent dans les zones d’étude, N’envisageant pas de déménager durant la période d’étude, Nationalité Malagasy, Signature de consentement éclairé (tuteur pour les mineurs avec assentiment)
Critères de non inclusion des femmes enceintes	De passage

Nature de données

Les données collectées et utilisées dans cette étude ont été de nature :

- Profils sociodémographiques (âge, niveau d’instruction, religion, situation matrimoniale, occupation, nombre d’habitant par ménage, etc.) des femmes enceintes
- Profils de santé de la femme
- Le recours et la perception sur la consultation prénatale
- La position du ménage par rapport aux centres de santé (distance entre le centre de santé le plus proche et le ménage)
- Le district et commune

2.3 Analyse de données

2.3.1 Outils

Le SGBD (Système de Gestion de Base de Données) utilisé a été le PostgreSQL. Les données prétraitées ont été enregistrées dans cette base.

Les traitements des données ont été faits sur le logiciel R version 3.4.0 avec l'éditeur Rstudio version 1.1.463 qui est multiplateforme et ayant une interface plus pratique et agréable [16].

R est un système qui est communément appelé langage et logiciel qui permet de réaliser des analyses statistiques (manipulation de données, calculs, etc). Sa conception a été fortement influencée par deux langages : S et Scheme. Il comporte un grand nombre de procédures statistiques telles que les modèles linéaires, les modèles linéaires généralisés [12], la régression non-linéaire, les séries chronologiques, les tests paramétriques et non paramétriques classiques, etc.

2.3.2 Préparation des données

Détection des doublons

Chaque individu ne doit pas figurer plus d'une fois dans la base.

Gestion des données manquantes

Les données manquantes ont été traitées selon les règles décrites par le tableau suivant :

TABLE 2.2 – Gestion des données manquantes

Variable qualitative	Variable quantitative
- Proportion inférieure à 10% : Imputation par la modalité la plus fréquente	- Proportion inférieure à 10% : Imputation simple par la moyenne ou médiane
- Proportion plus de 10% : Supprimer les observations correspondantes	- Proportion plus de 10% : Supprimer les observations correspondantes

Prévalence du recours aux CPN

La prévalence du recours aux consultations prénatales est calculée comme suit :

$$\text{Prévalence} = \frac{\text{Nombre de femmes ayant réalisé au moins une séance de CPN}}{\text{Nombre total des femmes de la population d'étude}} * 100$$

Recodage de la variable d'intérêt

Nous prenons en compte la recommandation de l'OMS qui recommande au moins 4 séances de CPN pour une femme enceinte durant sa grossesse.

Afin de bien analyser les niveaux du recours aux CPN chez les femmes, la variable dépendante prise est le **nombre de CPN faites** par une femme durant sa grossesse que nous avons codé en trois classes ordonnées dont :

1. **Aucune** : zéro(aucune) séance de CPN faite durant la grossesse
2. **Insuffisante** : 1 à 3 séances de CPN faites
3. **Complète** : 4 séances et plus

2.3.3 Analyses biostatistiques

Analyses descriptives

Les variables quantitatives ont été représentées par leur médiane avec les intervalles interquartiles respectifs tandis que les variables de type qualitatives ont été représentées par la proportion (en pourcentage) avec les intervalles de confiance à 95%.

Les tests statistiques utilisés ont été le test de χ^2 (Chi-deux) ou le test exact de Fisher selon leurs conditions d'applicabilité pour les variables qualitatives. Le test de Kruskal-Wallis (one-way analysis of variance) a été utilisé pour les variables quantitatives tout en fixant le seuil de significativité à 5%.

Analyses uni variées

Il s'agit de l'étude des associations entre chaque variables explicatives et la variable d'intérêt à l'aide de l'*odds - ratio* brut obtenu en appliquant la régression logistique ordinaire. Les variables retenues en analyses multivariées ont été celles dont le *p - value* est inférieur à 0,20 (suivant le protocole du projet EDRaS).

Analyses multivariées

Partant des variables sélectionnées en uni variées, les associations entre l'ensemble des variables explicatives et la variable dépendante ont été estimées à l'aide de l'*odds - ratio* ajusté.

Le modèle utilisé a été la régression logistique ordinaire [22] avec lequel nous avons appliqué deux méthodes, qui seront évalués, telles que la méthode à *LOGITS* adjacents et la méthode à *LOGITS* cumulés.

- La méthode des *LOGITS* adjacents permet d'estimer le *LOGIT* du passage d'un niveau à un autre qui lui est supérieur. C'est-à-dire la chance pour une femme de passer du niveau «aucune» en «insuffisante», ou de «insuffisante» en «complète».

Pour ce faire, le calcul se fait comme en multinomiale mais la modalité de référence change à chaque niveau. Écrivons nos trois niveaux comme : 1 =«Aucune», 2 =«Insuffisante» et 3 =«Compleète». Nous avons 3 – 1 = 2 équations du *LOGIT* à calculer. Par exemple pour le passage de 1 à 2, l'équation du *LOGIT* est :

$$LOGIT(\pi_2) = \ln\left(\frac{P(Y = 2|X)}{P(Y = 1|X)}\right) = \beta_{0,2} + \beta_{1,2}X_1 + \dots + \beta_{p,2}X_p \quad (2.1)$$

Pour chaque niveau k , le logarithme de l'*odds – ratio* est :

$$\ln(OR) = \ln\left(\frac{\frac{P(Y=k+1|X_i=x,\dots)}{P(Y=k|X_i=x,\dots)}}{\frac{P(Y=k+1|X_i=\bar{x},\dots)}{P(Y=k|X_i=\bar{x},\dots)}}\right) = \beta_{i,k}$$

Par suite, $OR = e^{\beta_{i,k}}$ est l'*odds – ratio* lié à la variable X : c'est l'*OR* d'effectuer au moins une séance de CPN au lieu de n'en faire aucune.

Pour faciliter l'interprétation des résultats, nous supposons que chaque variable agit de la même manière sur la variable dépendante quel que soit le niveau k , c'est-à-dire, $\beta_{i,k} = \beta_{i,k+1}$. Avec cette hypothèse, nous avons le modèle à pente parallèle :

$$\ln\left(\frac{P(Y = k + 1|X)}{P(Y = k|X)}\right) = \beta_{0,k} + \beta_1X_1 + \beta_2X_2 + \dots + \beta_pX_p \quad (2.2)$$

- Pour la méthode des *LOGITS* cumulatifs, le principe est de modéliser le *LOGIT* de dépasser un niveau donné. Plus précisément, elle permet d'estimer la probabilité de faire beaucoup plus (ou moins) de séances de CPN. Nous définissons la probabilité cumulée $P(Y \geq k|X) = \pi_k + \dots + \pi_3$ ($k = 1$ à 2).

L'équation des *LOGITS* cumulatifs se définit, en posant $k = 2$:

$$LOGIT(P(Y \geq 2|X)) = \ln\left(\frac{P(Y \geq 2|X)}{P(Y < 2|X)}\right) = \ln\left(\frac{\pi_2 + \pi_3}{\pi_1}\right) = \beta_{0,2} + \beta_{1,2}X_1, \beta_{2,2}X_2, \dots, \beta_{p,2}X_p \quad (2.3)$$

Avec le modèle à pente parallèle,

$$\ln\left(\frac{P(Y \geq 2|X)}{P(Y < 2|X)}\right) = \beta_{0,2} + \beta_1X_1 + \beta_2X_2 + \dots + \beta_pX_p \quad (2.4)$$

Par suite,

$$P(Y \geq 2|X) = \frac{e^{\beta_{0,2} + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p}}{1 + e^{\beta_{0,2} + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p}}$$

En général, pour avoir les $P(Y = k|X)$, on adopte la propriété suivante :

$$P(Y = k|X) = P(Y \geq k|X) - P(Y \geq k + 1|X)$$

- D'un autre côté, nous avons supposé que chaque district (voire même chaque commune) a sa propre tendance sur le recours aux CPN par rapport à celle de l'ensemble de la population. Les variables district et commune ont donc été utilisées comme variables à effets aléatoires vu qu'on voit de différence significative de proportion des femmes pratiquant la CPN entre ces lieux. Pour mettre en cause ces effets, nous avons associé au modèle ordinal à *LOGITS* cumulés l'effet mixte (aléatoire : le district puis la commune, et fixe : les autres variables explicatives). Les effets aléatoires ont été mesurés à l'aide de la constante (intercept) aléatoire.

L'effet district et commune ont été ainsi comparés par la méthode ANOVA. Cette méthode permet de comparer les critères d'information d'Akaike (ou Akaike Information Criterion : AIC) et le log-vraisemblance des deux modèles emboîtés. La significativité des différences entre les deux modèles est évaluée par le *p - value*.

Implémentation dans R

Pour implémenter ces modèles, nous avons utilisé les packages **ordinal** [8] et **VGAM** [21] qui existent sur R. Le package **ordinal** contient déjà des méthodes comme *clm()* pour faire appel à un modèle de régression ordinal à *LOGITS* cumulés et *clmm()* pour le modèle de régression ordinal à *LOGITS* cumulés et à effets mixtes. La méthode *vglm()* de **VGAM** a été utilisé pour le modèle de régression ordinal à *LOGITS* adjacents. D'autres packages ont été aussi utilisés selon notre besoin [7].

Évaluation des modèles : Performance et apport des résultats

Pour pouvoir comparer la performance des modèles utilisés, nous avons évalué leur capacité de prédiction. Pour ce, nous avons calculé les taux de mauvaises et bonne prédiction à l'aide de l'élaboration des matrices de confusion. Il s'agit d'un outil servant à mesurer la qualité d'un système de classification. Avec cette matrice, nous avons calculé le taux de mauvaises prédictions du modèle. Ce taux se calcule comme la proportion des valeurs prédites qui ne correspondent pas aux valeurs réelles.

Nous évaluons aussi nos modèles à partir des résultats qu'ils ont donnés. En effet, nous nous basons sur l'attente du projet EDRaS qui vise à détecter les variables qui expliquent

mieux les recours aux soins prénatals, et, en plus qui pourrait donner des idées pour des axes d'amélioration des qualités de soins dans les zones d'études.

Chapitre 3

RESULTATS, INTERPRETATIONS et DISCUSSIONS

3.1 Inclusion de la population

Le nombre de population incluse au premier et au deuxième passage n'a pas été le même. En effet, des femmes n'ont pas pu être retenues au deuxième passage de l'enquête pour différentes raisons.

Ci-après la figure montrant le flux d'inclusion des femmes durant notre étude :

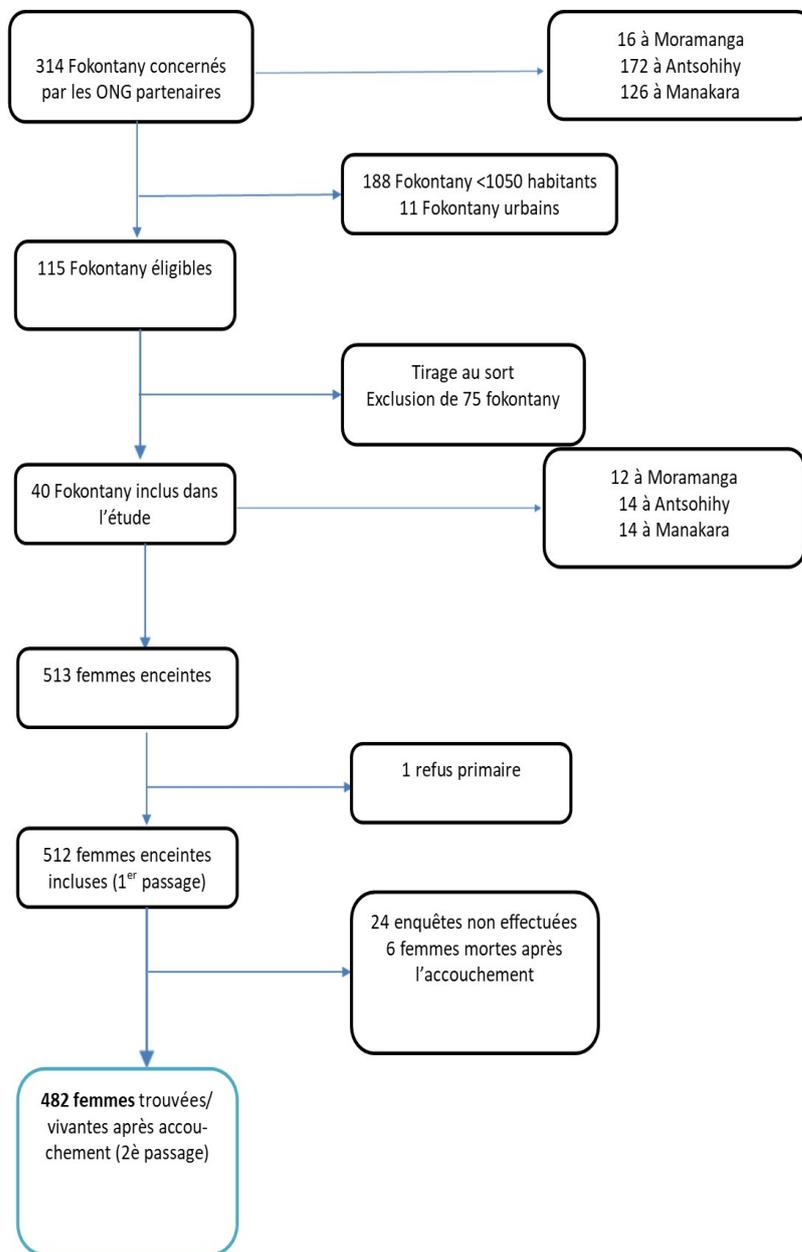


FIGURE 3.1 – Flux d’inclusion des femmes dans le cadre du projet EDRaS, IPM 2017-2018

3.2 Recours aux CPN

Concernant le nombre de séances de CPN effectuées par les femmes, la médiane a été de 3 ($IIQ = [2, 4]$) avec un minimum de 0 et un maximum de 9.

La figure qui suit représente la proportion des femmes par classe du recours aux soins prénatals.

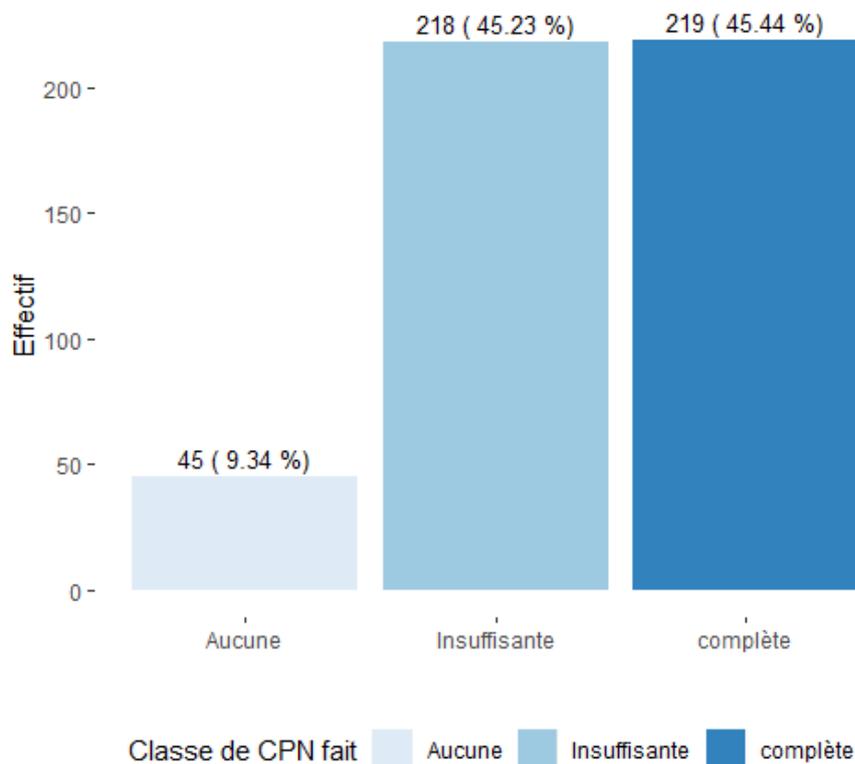


FIGURE 3.2 – Proportion des femmes par classe du recours aux CPN

Nous avons trouvé que 9,34% de la population d'étude n'ont effectué aucune séance de CPN durant leur grossesse. Nous constatons de même que les proportions des femmes en classe et complète sont presque identiques (218 et 219).

En se focalisant sur les districts de l'étude, nous constatons que la classe de recours aux CPN diffère significativement d'un district à l'autre ($p < 0,001$).

TABLE 3.1 – Association entre le nombre de CPN et les districts

Variable		Antsohihy N=156 (32,37%)	Moramanga N=164 (34,02%)	Manakara N=162 (33,61%)	p-value
Classe de recours aux CPN	Aucune	17 (10,90)	15 (9,15)	13 (8,02)	<0,001
	Insuffisante	110 (70,51)	63 (38,41)	45 (27,78)	
	Complète	29 (18,59)	86 (52,44)	104 (64,20)	

Les femmes de MANAKARA sont les plus probables à se faire compléter leur nombre de séance de CPN, vient ensuite les femmes de MORAMANGA. Nous constatons aussi que les femmes d'ANTSOHIHY ont presque (89,1%) commencé leur CPN mais la plupart n'ont pas atteint le nombre minimal recommandé par l'OMS.

Sur les communes, on voit que la commune d'AMBOHIBARY (district de MORAMANGA) domine les trois classes.

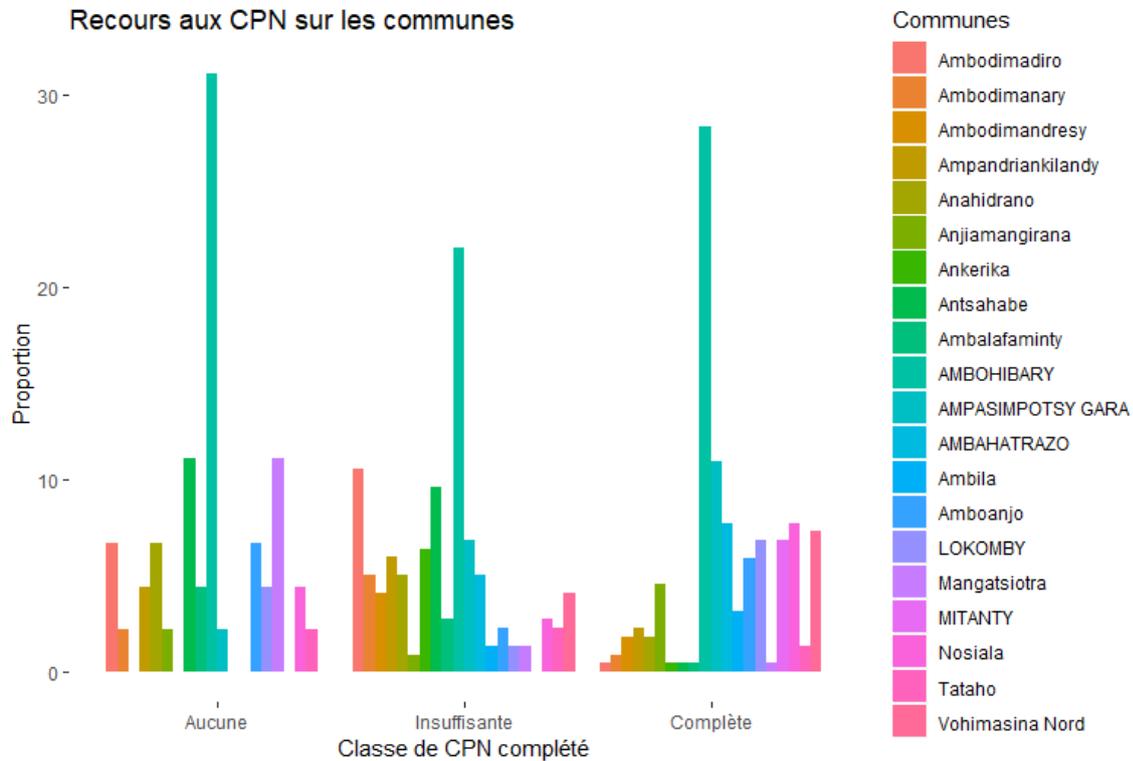


FIGURE 3.3 – Classe de Recours aux CPN par commune

3.3 Description de l'échantillon

Parmi les femmes de notre étude, il y avait encore des mineurs, et, la population la plus jeune est celle du district d'Antsohihy.

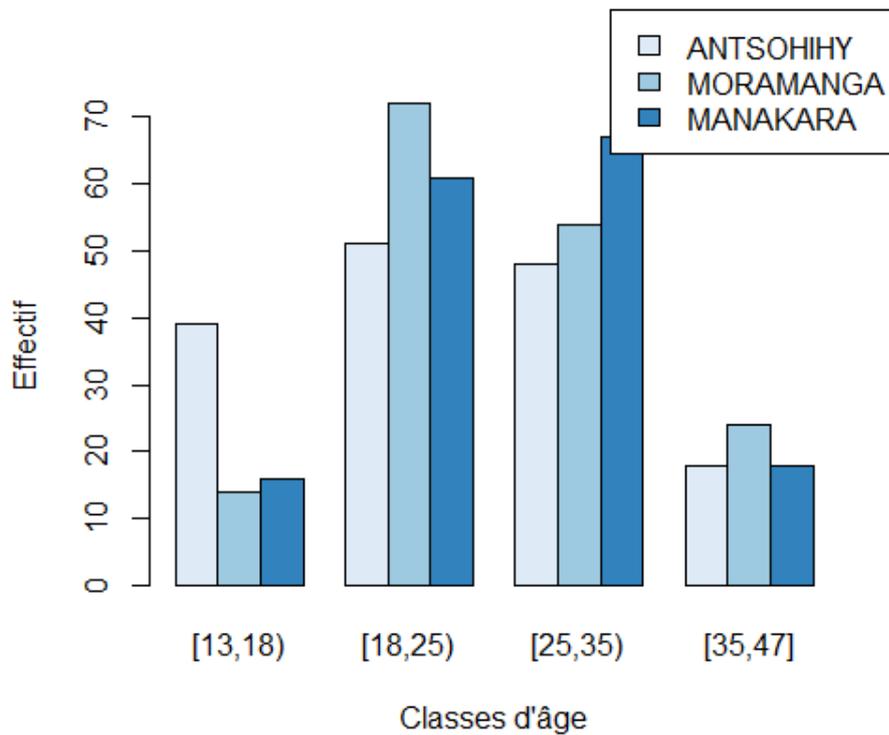


FIGURE 3.4 – Distribution des femmes dans chaque district selon les classes d'âge. Résultat EDRaS 2018

Les tableaux suivant nous montrent la comparaison des caractéristiques sociodémographiques, profil de santé et perception chez les femmes classées par leur nombre de séance (de CPN) complété.

TABLE 3.2 – Distributions des profils sociodémographiques selon les classes de recours aux CPN. Analyse descriptive - Résultat EDRaS 2018

		SEANCES DE CPN FAITES								
		TOTAL (n=482)		Aucune		Insuffisante		Complète		p-value
		n	%	n	%	n	%	n	%	
Age*		24	19-47	21	18-27	24,5	18-30	23,5	19-29	0,606
Instruction	Illettré	98	20,33	17	37,78	37	16,97	44	20,09	0,025
	Primaire	249	51,66	21	46,67	125	57,34	103	47,03	
	Secondaire	116	24,07	6	13,33	49	22,48	61	27,85	
	Supérieur	19	3,94	1	2,22	7	11	11	5,02	
Occupation	Non	77	15,98	7	15,56	28	12,84	42	19,18	0,195
	Oui	405	84,02	38	84,44	190	87,16	177	80,82	
Mariée	Non	89	18,46	10	22,22	52	23,85	27	12,33	0,006
	Oui	393	81,54	35	77,78	166	76,15	192	87,67	
Religion	Aucune	103	21,37	17	37,78	50	22,94	36	16,44	<0,001
	Musulman	58	12,03	4	8,89	43	19,72	11	5,02	
	Autre	60	12,45	2	4,44	22	10,09	36	16,44	
	Protestant	124	25,73	11	24,44	41	18,81	72	32,88	
	Catholique	137	28,42	11	24,44	62	28,44	64	29,22	

*Notons que la variable Age est ici présentée par la médiane et les intervalles interquartiles.

TABLE 3.3 – Distributions des profils de santé selon les classes de recours aux CPN. Analyse descriptive - Résultat EDRaS 2018

		SEANCES DE CPN FAITES								
		TOTAL (n=482)		Aucune		Insuffisante		Complète		p-value
		n	%	n	%	n	%	n	%	
Maladie	Non	403	83,61	37	82,22	172	78,90	194	88,58	0,022
	Oui	79	16,39	8	17,78	46	21,10	25	11,42	
Primigeste	Non	355	73,65	34	75,56	160	73,39	161	73,52	0,954
	Oui	127	26,35	11	24,44	58	26,61	58	26,48	
Grossesse souhaité	Non	61	12,66	12	26,67	31	14,22	18	8,22	0,002
	Oui	421	87,34	33	73,33	187	85,78	201	91,78	
Problème de santé durant la grossesse	Non	387	80,29	37	82,22	158	72,48	192	87,67	0,0003
	Oui	95	19,71	8	17,78	60	27,52	27	12,33	

TABLE 3.4 – Distributions des perceptions femmes et le rôle de leurs entourages selon les classes de recours aux CPN. Analyse descriptive - Résultat EDRaS 2018

											SEANCES DE CPN FAITES				
			TOTAL (n=482)		Aucune		Insuffisante		Complète		p-value				
			n	%	n	%	n	%	n	%					
Avis aux CPN	Pas pour		7	1,45	2	4,44	4	1,83	1	0,46					
	Pour		475	98,55	43	95,56	214	98,17	218	99,54					
Consultée par homme	Pas d'accord		76	15,77	9	20,00	29	13,30	38	17,35	0,364				
	D'accord		406	84,23	36	80,00	189	86,70	181	82,65					
INCITEE PAR :															
Entourage proche	Non		19	3,94	5	11,11	4	1,83	10	4,57	0,016				
	Oui		463	96,06	40	88,89	214	98,17	209	95,43					
Agent communautaire	Non		29	6,02	6	13,33	13	5,96	10	4,57	0,089				
	Oui		453	93,98	39	86,67	205	94,04	209	95,43					
Autorités locales	Non		165	34,23	20	44,44	66	30,28	79	36,07	0,140				
	Oui		317	65,77	25	55,56	152	69,72	140	63,93					
Utilisation passé de PF	Non		199	41,29	20	44,44	87	39,91	92	42,01	0,817				
	Oui		283	58,71	25	55,56	131	60,09	127	57,99					

3.4 Analyse des facteurs de recours aux CPN

3.4.1 Analyses uni variées

Les résultats en analyses uni variées sont représentés dans le tableau qui suit. Les *odds – ratio* brutes, les intervalles de confiance à 95% et la *p – value* décrivent les associations et la significativité. Rappelons que le seuil d’inclusion ici a été fixé à 0,20 et la méthode utilisée a été celle de *LOGITS* adjacents.

Nous avons onze (11) variables à inclure dans notre modèle comme le niveau d’étude, l’occupation, la situation matrimoniale, la religion, l’antécédent de maladie et l’état de santé actuel, le désir de la grossesse, le point de vue sur la CPN, le rôle de l’agent communautaire et la distance vers un centre de santé.

TABLE 3.5 – Association entre les profils sociodémographiques, profils de santé et la perception des femmes avec leurs recours aux soins prénatals. Analyses uni variées.

		OR	IC à 95%	p-value
Age	[13,18[1		
	[18,25[1,15	0,75-1,76	0,514
	[25,35[1,19	0,78-1,83	0,422
	[35,47[1,16	0,68-1,97	0,589
Instruction	Illettrée	1		
	Primaire	1,13	0,79-1,62	0,491
	Secondaire	1,63	1,07-2,49	0,024
	Supérieur	1,88	0,83-4,28	0,131
Occupation	Non	1		
	Oui	0,76	0,51-1,12	0,167
Mariée	Non	1		
	Oui	1,62	1,15-2,30	0,006
Religion	Aucune	1		
	Musulman	0,87	0,54-1,40	0,562
	Autre	2,66	1,55-4,56	<0,001
	Protestant	2,13	1,40-3,24	<0,001
	Catholique	1,61	1,09-2,40	0,017
Antécédent de maladie	Non			
	Oui	0,67	0,46-0,96	0,029

Première grossesse	Non			
	Oui	1,03	0,75-1,41	0,854
Grossese souhaitée	Non	1		
	Oui	1,99	1,33-2,98	<0,001
Problème de santé durant la grossesse	Non	1		
	Oui	0,63	0,45-0,88	0,007
Point de vue sur la CPN	Pas pour	1		
	Pour	3,06	1,02-9,20	0,047
Accepter si l'agent CPN est un homme	Non	1		
	Oui	0,94	0,64-1,38	0,762
INCITEE PAR : Entourage proche	Non	1		
	Oui	1,27	0,64-2,52	0,501
Agent communautaire	Non	1		
	Oui	1,71	0,98-2,98	0,058
Autorités locales	Non	1		
	Oui	1,01	0,76-1,35	0,933
Utilisation passée de PF	Non	1		
	Oui	1,00	0,75-1,32	0,982
Distance vers le CS le plus	0 à 5km	1		
	6 km et plus	0.61	0.42-0.88	<0.01

3.4.2 Analyses multivariées : Modèles obtenus

Partant des variables sélectionnées, nous avons utilisé les deux méthodes cumulés et adjacents.

Notons que les tableaux qui suivent présentent les résultats finals, c'est-à-dire après avoir construit le modèle initial avec les variables significatives en analyses uni variées et après des méthodes de sélection de variables pas à pas descendantes.

Méthodes des *LOGITS* adjacents

Le tableau suivant nous permet de voir en détail les *odds – ratio* du passage d'un niveau à un autre.

TABLE 3.6 – Facteurs favorisant le passage de niveau en niveau en nombre de séances de CPN effectuées par les femmes. *LOGITS* adjacents. Résultat EDRaS 2018

Classes/Niveaux de CPN fait				OR	IC à 95%	p-value
Grossesse désirée	Aucune	->	Non	1		
	Insuffisante		Oui	2.13	0.99-4.57	0.054
	Insuffisante		Non	1		
	-> Complète		Oui	1.95	1.05-3.63	0.035
Problème de santé durant la grossesse	Aucune	->	Non	1		
	Insuffisante		Oui	1.68	0.74-3.84	0.215
	Insuffisante		Non	1		
	-> Complète		Oui	0.36	0.22-0.60	<0.001

Interprétation : Nous avons trouvé deux variables significatives, qui sont le problème de santé durant la grossesse et le désir de la grossesse. En regardant en détail les *odds – ratio* du passage d'un niveau de CPN à un autre, nous remarquons que quand la grossesse est souhaitée, les femmes ont 1,95 fois plus de chances de se faire compléter le nombre de ses consultations plutôt que d'être dans la classe «insuffisante». D'autre part, nous voyons que les femmes ayant eu de problème de santé durant leurs grossesses ont moins de chances ($OR = 0,36$) d'aller au-delà de la classe «insuffisante» par rapport à celles qui n'ont pas eu de problème. Autrement dit, elles ont ($1/0,36=2,77$) fois plus risquées de rester dans la classe «insuffisante» plutôt que de se faire compléter leur nombre de consultation prénatale (classe «complète»).

Par contre, nous ne pouvons rien conclure sur le passage du niveau «aucune» au niveau «insuffisante».

Méthodes des *LOGITS* cumulés

Les tableaux suivants nous donnent les résultats utilisant la méthode des *LOGITS* cumulés, l'un utilisant la variable «districts» comme effet aléatoire et l'autre en utilisant la variable «commune» comme effet aléatoire.

1. Résultat en prenant le district comme effet aléatoire

TABLE 3.7 – Facteurs d’augmentation du nombre de séances de CPN chez les femmes enceintes sous l’effet aléatoire : District. *LOGIT* cumulé. Résultat EDRaS 2018

		OR	IC à 95%	p-value
Grossesse désirée	Non	1		
	Oui	1.97	1.10-3.52	<0.001
Problème de santé durant la grossesse	Non	1		
	Oui	0.58	0.36-0.91	0.03
Distance vers le CS le plus proche	0 à 5 km	1		
	6 km et plus	0.54	0.37-0.80	<0.01

Exponentielle de l’intercept(constante) aléatoire de l’enseble des districts : 1,55; de chaque district :

- ANTSOHIHY : 0,41
- MORAMANGA : 1,29
- MANAKARA : 1,86

2. Résultat en prenant la commune comme effet aléatoire

TABLE 3.8 – Facteur d’augmentation du nombre de séances de CPN chez les femmes enceintes sous l’effet aléatoire : Commune. *LOGIT* cumulé. Résultat EDRaS 2018

		OR	IC à 95%	p-value
Grossesse désirée	Non	1		
	Oui	2.49	1.42-4.37	<0.01

Exponentielle de l’intercept aléatoire de toute la population d’étude : 1,42, de chaque commune :

District d’ANTSOHIHY : Ambodimadiro= 0,29; Ambodimanary= 0,45; Ambodimandresy= 0,88; Ampandriankilandy=0,49; Anahidrano= 0,43; Anjiamangirana= 3,37; Ankerika= 0,37; Antsahabe= 0,25; Ambalafaminty= 0,34;

District de MORAMANGA : Ambohibary= 1,14; Ampasipotsoy Gara= 1,97;

District de MANAKARA : Ambahatrazo= 2,20; Ambila= 2,63; Amboanjo= 1,54; Lokomby= 2,67; Mangarsiotra= 0,12; Mitanty= 11,12; Nosiala= 2,38; Tataho= 0,71; Vohimasina Nord=2,28

Interprétations : En bref, ces tableaux nous donnent que le désir de la grossesse (de manière bénéfique), l'existence de problème de santé durant la grossesse et le fait d'habiter à plus de 5 km du CSB (de manière nocive) agissent sur les facteurs de recours aux consultations prénatales. Globalement, les effets aléatoires ont montré que, quand les variables explicatives (effets fixes) valent zéro, il y a en moyenne 1,5 fois de chances de se faire évoluer le recours aux CPN, mais quand nous projetons cela sur chaque district, on voit que les femmes de MANAKARA et de MORAMANGA surpassent cette chance (respectivement de 1,86 et de 1,29 fois), tandis que la tendance pour ANTISOHIHY décale en dessous de cette moyenne (0,41). De même sur les communes, on remarque que les femmes de la commune d'Anjiamangirana surpassent largement la moyenne (de chance de recours aux CPN).

3.5 Évaluation des méthodes

3.5.1 LOGITS Adjacents et Cumulés

Qualité des résultats

En résumé, voici les résultats que les deux méthodes ont donnés :

TABLE 3.9 – Tableau comparatif des résultats des deux méthodes du modèle logistique ordinaire

Facteurs trouvés	LOGITS Adjacents	LOGITS Cumulés
Désir du grossesse	Odds – ratio = 1,95 (se faire compléter ses séances de CPN au lieu de rester en insuffisante)	Odds – ratio_{districts} = 1,97 Odds – ratio_{communes} = 2,49 (faire au moins une séance de CPN ou se faire compléter ses CPN)
Problème de santé	Odds – ratio = 0,36 (se faire compléter ses séances de CPN au lieu de rester en insuffisante)	Odds – ratio_{districts} = 0,58 (faire au moins une séance de CPN ou se faire compléter ses CPN)
Distance vers le CSB		Odds – ratio_{districts} = 0,54 (faire au moins une séance de CPN ou se faire compléter ses CPN)

La méthode des *LOGITS* cumulés s'est montré plus informative. En effet, nous avons obtenu une variable de plus qui est la distance vers le centre de santé le plus proche comme facteur défavorisant le recours aux CPN. Or, cette variable permet de proposer des axes d'amélioration des stratégies visant à augmenter l'accès et le taux de recours aux soins, tel est le genre résultat attendu du projet EDRaS.

De plus, avec la méthode à *LOGITS* adjacents, chaque *odds – ratio* est propre à deux niveaux adjacents fixés. Par cette méthode, il se peut que des passages en niveaux différents ne soient pas interprétables (comme notre cas ici, nous ne pouvons rien conclure sur le passage du niveau «insuffisante» en «complète». Quant à elle, la méthode à *LOGITS* cumulatifs est plus large. Son rapport des côtes s’interprète comme le passage en un niveau supérieur (ou inférieur) quel que soit le niveau actuel.

Performance des modèles

Voici les matrices de confusion nous permettant d’évaluer la proportion des mauvaises et bonnes prédictions :

TABLE 3.10 – Matrice de confusion obtenu par le modèle à *LOGITS* adjacents

		Classe réelle		
		Aucune	Insuffisante	Complète
Classe prédite	Aucune	1	1	0
	Insuffisante	23	125	46
	Complète	21	92	173

Taux de mauvais classement : 37,97% ($21 + 23 + 1 + 92 + 46 = 183$ sur 482).

TABLE 3.11 – Matrice de confusion obtenu par le modèle à *LOGITS* cumulés

		Classe réelle		
		Aucune	Insuffisante	Complète
Classe prédite	Aucune	3	0	0
	Insuffisante	21	131	47
	Complète	21	87	172

Taux de mauvais classement : 36,51% ($21 + 21 + 87 + 47 = 176$ sur 482).

Ces tableaux nous donnent la précision des modèles. Nous voyons que le taux de mauvaises prédictions du modèle à *LOGITS* adjacents est supérieur à celui du modèle cumulé (37,97% > 36,51%).

Encore, la méthode à *LOGITS* cumulés prédit mieux notre variable d'intérêt que la méthode à *LOGITS* adjacents.

Malgré ces différences, l'interprétation ensemble des deux méthodes enrichit notre étude sur les déterminants du recours aux soins prénatals.

3.5.2 Modèle à *LOGITS* Cumulés : Comparaison des deux résultats avec effets aléatoires districts et communes

En comparant les deux modèles avec ANOVA, nous avons trouvé un $p - value = 1$. Le AIC (Akaike Information Criterium) pour le modèle à effet districts est de 847,29 et celui à effet communes est de 839,77.

Ainsi, nous en déduisons que les modèles en utilisant les districts et les communes comme effet aléatoire ne se diffèrent pas significativement.

Conclusion

Cette étude a permis de répondre aux objectifs fixés dont l'estimation de la prévalence du recours aux CPN dans trois districts (ANTSOHIHY, MORAMANGA et MANAKARA) et l'application du modèle logistique ordinal à effets mixtes afin de déterminer les facteurs influençant ce recours aux CPN en se basant sur la norme recommandée par l'OMS.

Le modèle à *LOGITS* cumulés a été trouvé plus informatif et performant par rapport au modèle à *LOGITS* adjacents. Malgré ces faits et la différence d'informations apportées par les modèles, leurs résultats peuvent être interprétés ensemble. De plus, les deux modèles à effet mixte (districts et communes) n'ont pas été significativement différents.

Nous avons trouvé que la majorité des femmes ont effectué au moins une séance de CPN durant leur grossesse. Trois facteurs déterminants (favorisant ou défavorisant le recours) ont été trouvés après avoir appliqué les deux modèles susmentionnés. Le désir de la grossesse augmente les chances qu'une femme aille se faire compléter ses CPN. Par contre, le fait d'habiter loin (plus de 5 km) du centre de santé et le fait d'avoir un problème de santé durant la grossesse diminuent les chances pour les femmes d'aller faire plus de CPN. Aussi, les femmes ayant eu de problème de santé durant la grossesse sont moins probables de dépasser le nombre minimal de CPN recommandé. Si on ne tient pas compte de l'effet des autres variables, les femmes de chaque district (et de chaque commune de résidence) a sa tendance spécifique par rapport à la tendance globale sur la complétude de ses séances de CPN.

Ce travail a aidé à l'atteinte de l'objectif du projet EDRaS. En effet, les variables que nous avons obtenues permettent de donner des idées à l'amélioration de l'accès aux soins dans ces trois districts. Pour que toutes les grossesses soient désirées, nous suggérons de renforcer la sensibilisation des femmes à l'utilisation des méthodes contraceptives et surtout de les former pour bien comprendre le planning familial. Nous proposons aussi de multiplier les centres de santé, et même dans les lieux en difficulté des moyens de transport.

Cette étude ne nous a pas seulement permis d'apprendre de nouvelles méthodes de la régression logistique ordinaire à effet mixte, mais elle a aussi permis de se familiariser dans le monde de travail et surtout de s'intégrer dans un nouvel environnement tel que l'analyse des données de santé.

Cependant, la caractéristique de notre donnée d'étude n'a pas été très adaptée pour appliquer le modèle de régression ordinaire. En effet, la proportion des femmes dans chaque classe (Aucune, insuffisante, complète) n'a pas été proportionnelle. Les résultats pourront être plus fiables si les données correspondent aux hypothèses du modèle choisi.

Aussi, une limite en santé publique, l'étude s'est uniquement focalisée sur le nombre de séances de CPN effectuées par les femmes durant la grossesse. Mais nous n'avons pas tenu compte ni de la date du début des consultations ni de l'intervalle entre les séances. Nous n'avons donc

pas pu évaluer le retard des débuts de consultation ainsi que la régularité des intervalles entre les différentes consultations. Il serait ainsi intéressant de considérer ces variables dans une étude ultérieure plus approfondie.

Cette étude pourra aussi être étendue sur d'autres districts de Madagascar en apportant des améliorations sur les modèles que nous avons trouvés mais appliqués sur un plus grand jeu de données.

Bibliographie

- [1] ADELIN, G. *et al.* *Principaux modèles utilisés en régression logistique*. 2011.
- [2] ANTOINE, B. *et al.* *Modèles à effets aléatoires pour l'analyse de la Qualité de Vie en Oncologie*. juin 2014.
- [3] BENOÎT-PAUL, H. *Régression avec une variable dépendante ordinale : comparaison de la performance de deux modèles logistiques ordinaires et du modèle linéaire classique à l'aide de données simulées*. PhD thesis, Université Laval Québec, 1998.
- [4] BENZAKI, Y. Multivariate regression. *Mr. Mint Machine Learning made easy* (2017).
- [5] BOUYER, J. *et al.* *Epidémiologie - Principes et méthodes quantitatives*. 1995.
- [6] CATHERINE, A. Le modèle linéaire à effets mixtes pour analyser des données génétiques provenant de familles, 2015.
- [7] CHONGSUVIVATWONG, V. *epicalc : Epidemiological calculator*, 2012. R package version 2.15.1.0.
- [8] CHRISTENSEN, R. H. B. ordinal—regression models for ordinal data, 2018. R package version 2018.8-25. <http://www.cran.r-project.org/package=ordinal/>.
- [9] DES STATISTIQUES SANITAIRES, S. *ANNUAIRE DES STATISTIQUES DU SECTEUR SANTE*. 2013.
- [10] INSTAT. *Enquête nationale sur le suivi des Objectifs du Millénaire pour le Développement à Madagascar : OBJECTIF 05*. 2013.
- [11] JEAN, B. La régression logistique en épidémiologie. *M2 recherche en santé publique, Université Paris V, XI, XII* (2012).
- [12] LAURENT, R. *Régression logistique avec R*.
- [13] MICHAËL, G. *Régression logistique*.
- [14] MILLOT, G. *Comprendre et réaliser les tests statistiques à l'aide de R*. septembre 2011.
- [15] PETTAN, D., AND MESBAH. Cumulative or adjacent logits : which choice an ordinal logistic latent variable model?
- [16] R CORE TEAM. *R : A Language and environment for statistical computing*. R Foundation for statistical computing, Vienna, Austria, 2018.
- [17] RAKOTOMALALA, R. *Régression logistique polytomique*.
- [18] RAKOTOMALALA, R. *Pratique de la régression logistique : Régression logistique binaire et polytomique*. mai 2017.

- [19] RAKOTOSEHENO, H., AND RAKOTONIRINA, E. Consultations prénatales et mortalité périnatale à madagascar. *Journal de Gynécologie Obstétrique et Biologie de la Reproduction*, Vol. 37 (2008), 505–509.
- [20] THOMAS YEE, W. Vgam family function for categorical data. *Department of Statistics, University of Auckland, Beta version 0.5-15* (2004).
- [21] THOMAS YEE, W. Two-parameter reduced-rank vector generalized linear models, 2013. *Computational Statistics and Data Analysis*. <http://ees.elsevier.com/csda/>.
- [22] TRAISSAC, P. et al. régression logistique vs autres modèles linéaires généralisées pour l'estimation de rapports de prévalences. *Rev. Epidém. et Santé Publ, Masson Paris* (1999).

Annexe 1

Définitions

Taux de mortalité maternelle : Indice statistique de la mort maternelle qui exprime le rapport entre le nombre de femmes décédées, à la suite de conséquences obstétricales directes ou indirectes, pendant leur grossesse ou lors des 42 jours après l'accouchement, et le nombre de naissances vivantes.

Critère d'information d'Akaike(AIC) : Une mesure de la qualité d'un modèle statistique (proposée par Hirotugu Akaike en 1973).

Loi multinomiale : Une généralisation de la loi binomiale, qui concerne le nombre de succès dans n épreuves de Bernoulli indépendantes donnant chacune un résultat binaire.

Modèles emboîtés : Deux modèles sont dits "emboîtés" si l'un peut être considéré comme un cas particulier de l'autre. Cela revient à comparer un modèle de référence à un modèle réduit. C'est-à-dire comparer deux modèles dont les variables utilisées dans l'un est compris dans l'autre.

Annexe 2

Grossesses indésirables

	Ensemble		Antsohihy		Moramanga		Manakara	
	N	(%)	N	(%)	N	(%)	N	(%)
Oui	34	6,6	11	6,5	11	6,4	12	7,0
Non	339	66,2	101	59,4	119	69,6	119	69,6
Non applicable NA	139	27,2	58	34,1	41	24,0	40	23,4
Total	512	100	170	100	171	100	171	100
Nombre de grossesses non désirées								
Moyenne	2		2		2		1	
Minimum	1		1		1		1	
Maximum	6		3		6		2	
Acte suite à la grossesse non désirée								
J'ai gardé	27	79,4	5	45,4	11	100	11	91,7
J'ai avorté	7	20,6	6	54,6	-		1	8,3
Total	34	100	11	100	11	100	12	100

FIGURE 3.5 – Grossesse non désirée. Source Projet EDRaS- IPM 2017-2018

Annexe 3

Inclusion des Fokontany

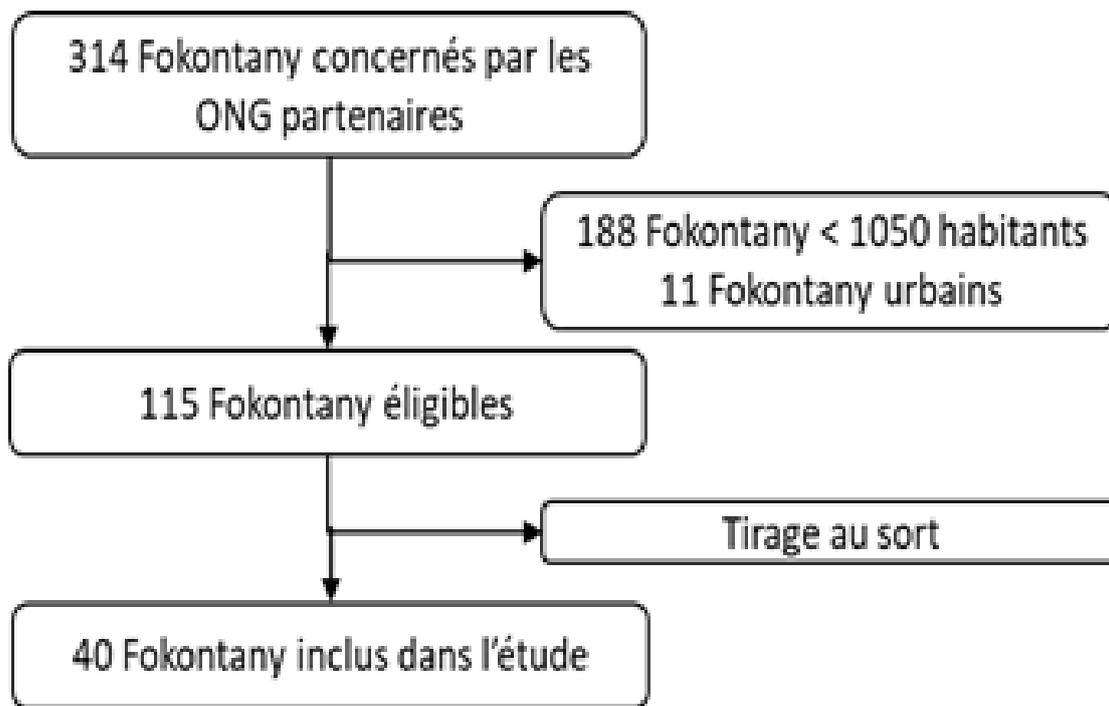


FIGURE 3.6 – Diagramme de flux d'inclusion des Fokontany. Source Projet EDRaS- IPM 2017-2018



FIGURE 3.7 – Carte des Fokontany inclus dans le district de Moramanga. Source Projet EDRaS-IPM 2017-2018

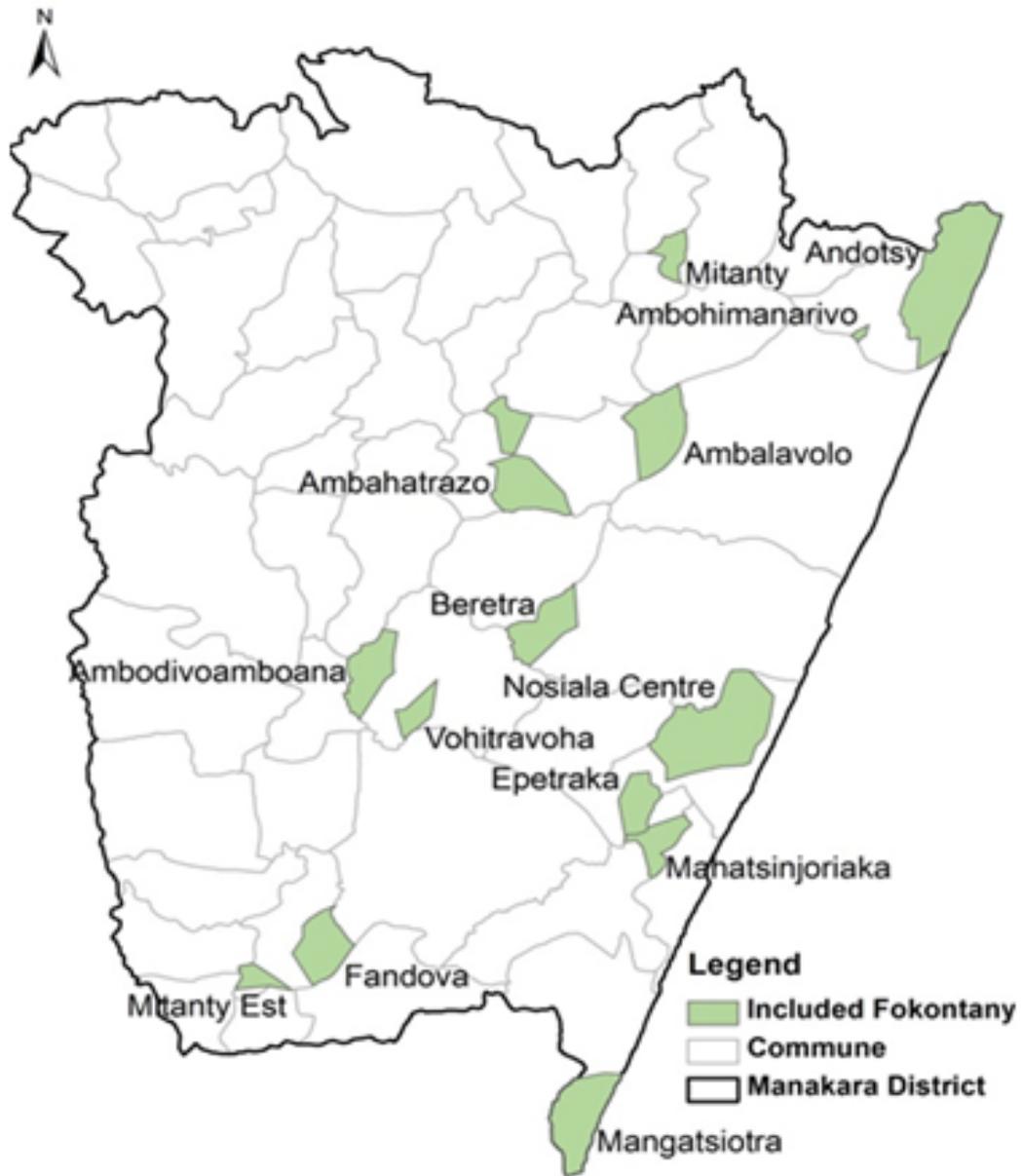


FIGURE 3.8 – Carte des Fokontany inclus dans le district de Manakara. Source Projet EDRaS-IPM 2017-2018

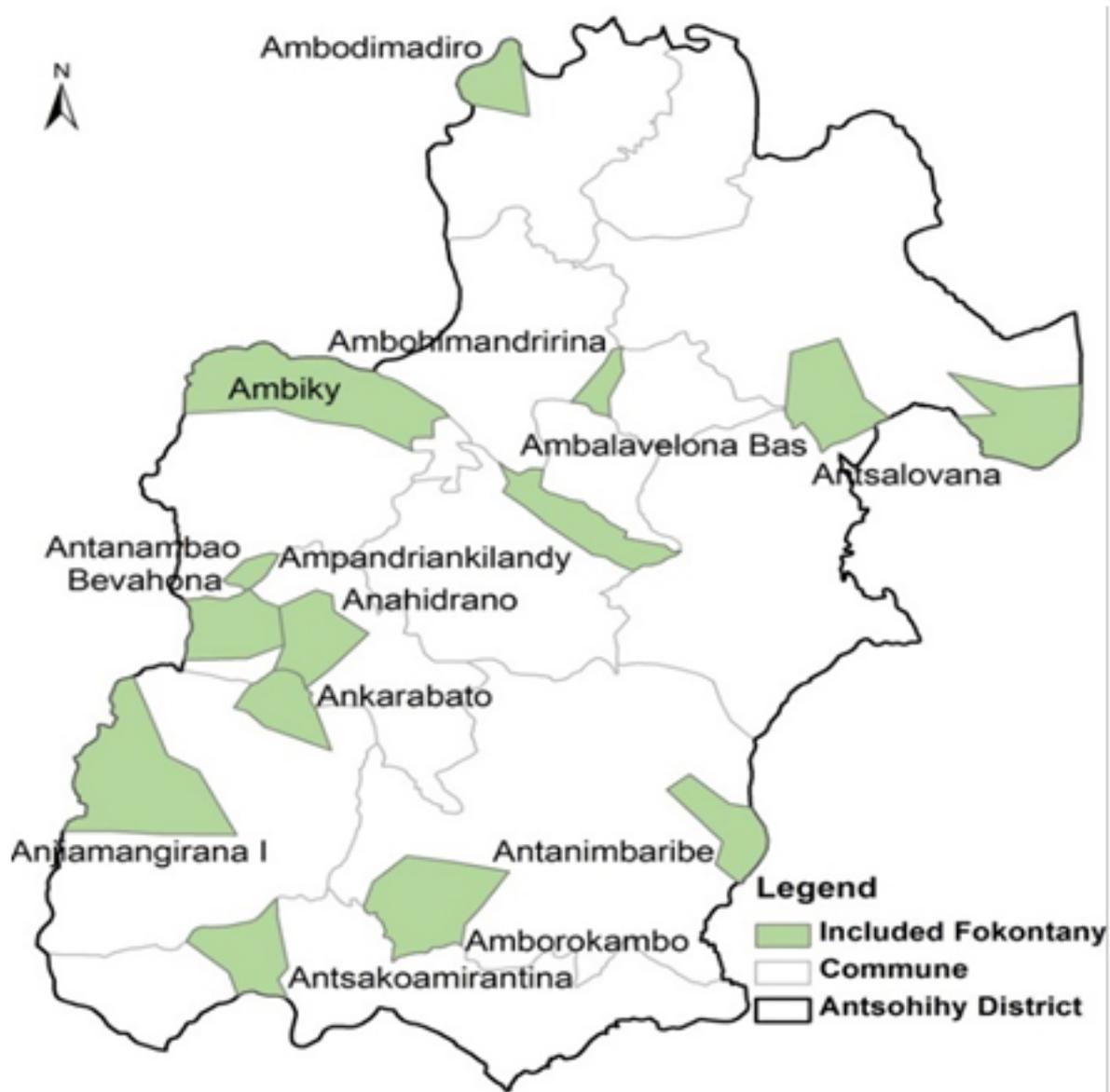


FIGURE 3.9 – Carte des Fokontany inclus dans le district d’Antsohihy. Source Projet ED RaS-IPM 2017-2018

Auteur : RAHAINGOVAHOAKA Nirintsoa Fara
033 18 383 17 – nirintsoafara@gmail.com

**Facteurs de réalisation des consultations prénatales dans trois districts de Madagascar :
APPLICATION DU MODELE DE REGRESSION POLYTOMIQUE ORDINALE A EFFET
MIXTE**

Résumé

La mortalité et la morbidité maternelle et infantile restent élevées à Madagascar. La pratique des Consultations Prénatales (CPN) constitue un élément clé dans la réduction de ces risques. Selon la norme de l'Organisation Mondiale de la Santé (OMS), au moins quatre séances de CPN sont recommandées pendant la grossesse. Il est essentiel de déterminer les facteurs qui influencent la réalisation de ces consultations dans le pays. Cette étude permet d'analyser ces facteurs en utilisant des modèles de régression logistique ordinaire dans les districts d'Antsohihy, de Moramanga et de Manakara. Des analyses uni variées ont été effectuées pour inclure les variables explicatives dans les modèles multivariés. Deux méthodes des *LOGITS* cumulés à effet mixte et des *LOGITS* adjacents ont été utilisées et le modèle à *LOGITS* cumulés a été plus informatif avec plus de variables significatives. Les effets aléatoires des districts et des communes de résidence ont été considérés. La prévalence du recours aux soins prénatals a été de 90,67%. Nos modèles ont montré que la distance vers le centre de santé ($OR=0,54/IC=[0,37-0,80]$) et le problème de santé durant la grossesse ($OR=0,58/IC=[0,36-0,91]$) ont été des facteurs de non recours aux consultations prénatals, contrairement au désir de la grossesse ($OR=1,97/IC=[1,10-3,52]$). Par conséquent, le renforcement du réseau de santé communautaire en milieu rural pourrait être une solution pour améliorer la santé materno-infantile. De plus, une étude élargie sur d'autres districts de Madagascar serait nécessaire pour confirmer nos résultats.

Mots clés : Santé materno-infantile, Consultations Prénatales, modèle logistique ordinal, modèle à effet mixte.

Abstract

Maternal and child mortality remains high in Madagascar. The practice of AnteNatal Care (ANC) is a key element in reducing these risks. According to the World Health Organization (WHO) standard, four sessions of ANC are at least required during pregnancy. It is necessary to determine the factors that influence the completion of these consultations. This study analyzes these factors using ordinal logistic regression models setting in ANTISOHIHY, MORAMANGA and MANAKARA. Univariate analyses were performed to include explanatory variables into multivariate models. Two variant of *LOGITS* methods such as cumulative with mixed-effect and adjacent have been used, and the cumulative model was more informative with more significant variables. Random effects on both district and the commune of residence levels were considered. ANC prevalence was 90.67%. Our models showed that the distance to the health center ($OR = 0.54/CI = [0.37 - 0.80]$) and the health matter during pregnancy ($OR = 0,58/CI = [0,36 - 0.91]$) were risk factors of prenatal cares, in contradiction of the pregnancy desire ($OR = 1.97/CI = [1.10 - 3.52]$). Therefore, strengthening the community health network in rural areas could be a solution to improve mother and child health. In addition, an extended study in other districts of Madagascar would be useful to confirm our results.

Key words : Mother and Child Health, Antenatal Consultations, ordinal logistic models, mixed-effect models

Rapporteur : Dr ANDRIAMAROZAKANIAINA Tahiry
Maître de Conférences